

2AP

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
23 December 2004 (23.12.2004)

PCT

(10) International Publication Number  
**WO 2004/111197 A2**

- (51) International Patent Classification<sup>7</sup>: C12N
- (21) International Application Number: PCT/US2004/018492
- (22) International Filing Date: 10 June 2004 (10.06.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- |            |                             |    |
|------------|-----------------------------|----|
| 60/477,218 | 10 June 2003 (10.06.2003)   | US |
| 60/483,387 | 27 June 2003 (27.06.2003)   | US |
| 60/497,599 | 25 August 2003 (25.08.2003) | US |
- (71) Applicants (for all designated States except US):  
TRUSTEES OF BOSTON UNIVERSITY [US/US];  
One Sherborn Street, Boston, MA 02215 (US).  
AFFYMETRIX, INC. [US/US]; 3380 Central Ex-  
pressway, Santa Clara, CA 95051 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): BRODY, Jerome,  
S. [US/US]; 22 Laudholm Road, Newton, MA 02458  
(US). SPIRA, Avrum [CA/US]; 1105 Massachusetts  
Avenue, Apt.6G, Cambridge, MA 02138 (US). SHAH,  
Nila [US/US]; 21555 Addington Court, Cupertino, CA  
95014 (US). PALMA, John, F. [US/US]; 2567 CraneFord  
Way, San Ramon, CA 94583 (US).
- (74) Agents: PIERCE, Scott, N. et al.; Hamilton, Brook, Smith  
& Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Con-  
cord, MA 01742-9133 (US).
- (81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,  
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,  
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,  
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,  
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,  
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,  
ZW.
- (84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,  
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,  
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,  
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, ML, MR, NE, SN, TD, TG).
- Published:  
— without international search report and to be republished  
upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 2004/111197 A2

(54) Title: GENE EXPRESSION SIGNATURES, METHODS AND COMPOSITIONS FOR DIAGNOSING DISORDERS OF THE LUNG

(57) Abstract: A minimally invasive sample procurement method for obtaining airway epithelial cell RNA that can be analyzed by expression profiling, e.g., by array-based gene expression profiling, is disclosed. These methods can be used to identify patterns of gene expression that are diagnostic of lung disorders, e.g., cancer, to identify subjects at risk for developing lung disorders and to custom design an array, e.g., a microarray, for the diagnosis or prediction of lung disorders or susceptibility to lung disorders. Arrays and informative genes are also disclosed for this purpose.

-1-

GENE EXPRESSION SIGNATURES, METHODS AND COMPOSITIONS FOR  
DIAGNOSING DISORDERS OF THE LUNG

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No.  
5 60/477,218, filed June 10, 2003, U.S. Provisional Application No. 60/483,387, filed  
June 27, 2003, and U.S. Provisional Application No. 60/497,599, filed August 25,  
2003.

The entire teachings of the above applications are incorporated herein by  
reference.

10 GOVERNMENT SUPPORT

The invention was supported, in whole or in part, by grant ES00354 from  
The Katherine and Roland Douglass Fund (American Cancer Society) and by grant  
HL07035 from the National Institute of Health. The Government has certain rights  
in the invention.

15 BACKGROUND OF THE INVENTION

Lung cancer claims more than 150,000 lives every year in the United States,  
exceeding the combined mortality from breast, prostate and colorectal cancers.  
Cigarette smoking is the most predominant cause of lung cancer. Presently, 25% of  
the U.S. population smokes, but only 10% to 15% of heavy smokers develop lung  
20 cancer. Former smokers remain at risk for developing cancer and now constitute a  
large reservoir of new lung cancer cases. Approximately 85% of all subjects with  
lung cancer die within three years of diagnosis. Unfortunately survival rates have  
not changed substantially in the past several decades, in large part because there are  
no effective methods for identifying smokers who are at highest risk for developing  
25 lung cancer and no effective tools for early diagnosis.

-2-

One major hurdle in developing an early detection screen for lung cancer is that present methods for diagnosis require removal of tissue from inside the lung. Moreover, while it appears that a subset of smokers are more susceptible to the carcinogenic effects of cigarette smoke and are more likely to develop lung cancer, particular risk factors, and particularly genetic risk factors, for individuals have gone largely unidentified.

#### SUMMARY OF THE INVENTION

Work described herein demonstrates the utility of a minimally invasive sample procurement method and gene expression-based tools for the diagnosis of diseases of the lung, particularly lung cancer. Work described herein is further based on the identification of unique sets of expressed genes associated with smokers and non-smokers that constitute expression signatures. Epithelial cell gene expression profiles obtained from relatively accessible sites can provide important diagnostic and therapeutic information which can be applied to diagnose and treat lung disorders. For example, the gene expression profiles or signatures disclosed herein can be used to distinguish between, e.g., non-smokers and smokers, smokers and smokers with cancer, non-smokers and smokers with cancer, or a combination thereof.

In particular, work described herein provides a minimally invasive sample procurement method for obtaining airway epithelial cell RNA that can be analyzed by expression profiling, e.g., by array-based gene expression profiling. These methods can be used to determine how airway epithelial cell gene expression profiles are affected by cigarette smoke and how these profiles differ in smokers with and without lung cancer. These methods can also be used to identify patterns of gene expression that are diagnostic of lung disorders, e.g., cancer, and to identify subjects at risk for developing lung disorders. All or a subset of the genes identified according to the methods described herein (e.g., the genes disclosed in FIGS. 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E) can be used to design an array, e.g., a microarray, specifically intended for the diagnosis or prediction of lung disorders or susceptibility to lung disorders, optionally in combination with

probes intended for the diagnosis or prediction of other disorders (e.g., in combination with probes intended for the diagnosis or prediction of other types of cancer), based on the gene expression signatures exhibited by smokers and non-smokers as shown in FIGS. 6A-6R. For example, an array can be designed which  
5 has probes for one or more of the genes disclosed herein immobilized thereon. The efficacy of such custom-designed arrays can be further tested, for example, in a large clinical trial of smokers.

In one embodiment, the invention relates to a method of distinguishing a biological sample from a non-smoker and a biological sample from a smoker, and/or  
10 a biological sample from a smoker and a biological sample from a smoker with cancer, and/or a biological sample from a non-smoker and a biological sample from a smoker with cancer, wherein the method comprises obtaining a biological sample from an individual to be diagnosed; determining the level of expression of one or more informative genes in said sample, and comparing the determined level of  
15 expression with the corresponding level of expression in an appropriate control (e.g., one or more of a non-smoker, a smoker and a smoker with cancer). In a preferred embodiment, the biological sample is obtained from an airway epithelial cell. In one embodiment, the airway epithelial cell is obtained from a bronchoscopy or buccal mucosal scraping.

20 In a preferred embodiment the informative genes are selected from the group consisting of the genes in FIGS. 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E. Preferably the expressions of two or more, five or more, ten or more, fifteen or more, twenty or more, fifty or more, one hundred or more, or one thousand or more informative genes are determined.

25 In a particular embodiment, the biological sample is a nucleic acid sample, and in a preferred embodiment the level of expression of one or more informative genes is determined using oligonucleotides which hybridize to the one or more informative genes. In one embodiment the oligonucleotides are immobilized on a substrate. In a particular embodiment, the nucleic acid sample is RNA.

30 In another particular embodiment, the biological sample comprises a gene expression product, and in a preferred embodiment the level of expression of one or



more informative genes is determined using polymers (e.g., polypeptides) which bind to gene expression products encoded by the one or more informative genes. In a particular embodiment the polymers are immobilized on a substrate.

In one embodiment, the invention relates to a method of diagnosing a disease  
5 of the lung, in particular lung cancer, comprising obtaining a biological sample from an individual to be diagnosed; and determining the expression of one or more informative genes in said sample, wherein increased expression of an informative gene whose expression is increased in individuals having a disease of the lung, or decreased expression of an informative gene whose expression is decreased in  
10 individuals having a disease of the lung, is indicative of a disease of the lung in the individual. In one embodiment, individual to be diagnosed is an individual who has been exposed to tobacco smoke, an individual who has smoked, or an individual who smokes. In a preferred embodiment, the biological sample is obtained from an airway epithelial cell. In one embodiment, the airway epithelial cell is obtained from  
15 a bronchoscopy or buccal mucosal scraping.

In a preferred embodiment of the method, the informative genes are selected from the group consisting of the genes shown in FIGS. 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E. Preferably the expressions of two or more, five or more, ten or more, fifteen or more, twenty or more, fifty or more, one  
20 hundred or more, or one thousand or more informative genes are determined.

In one embodiment, the disease of the lung is selected from the group consisting of asthma, chronic bronchitis, emphysema, primary pulmonary hypertension, acute respiratory distress syndrome, hypersensitivity pneumonitis, eosinophilic pneumonia, persistent fungal infection, pulmonary fibrosis, systemic  
25 sclerosis, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, and lung cancer, such as adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, and benign neoplasms of the lung (e.g., bronchial adenomas and hamartomas).

In a particular embodiment, the biological sample is a nucleic acid sample,  
30 and in a preferred embodiment the level of expression of one or more informative genes is determined using oligonucleotides which hybridize to the one or more

informative genes. In one embodiment the oligonucleotides are immobilized on a substrate. In a particular embodiment, the biological sample is RNA. In a preferred embodiment, the expression is determined using a microarray having one or more probes (e.g., oligonucleotides) for said one or more genes immobilized thereon.

5 In another particular embodiment, the biological sample comprises a gene expression product, and in a preferred embodiment the level of expression of one or more informative genes is determined using polymers (e.g., polypeptides) which bind to gene expression products encoded by the one or more informative genes. In a particular embodiment the polymers are immobilized on a substrate.

10 The invention further relates to a method of obtaining a biological sample for use in expression analysis comprising obtaining an airway epithelial cell sample from an individual. The method may further comprise rendering nucleic acid molecules in said cell sample available for hybridization, and/or rendering gene expression products in said cell sample available for binding to polymers.

15 The invention also relates to a method of treating a disease of the lung, in particular lung cancer, comprising administering to an individual in need thereof an effective amount of an agent which increases the expression of an informative gene whose expression is decreased in said individual as compared with a normal individual.

20 The invention further relates to a method of treating a disease of the lung, in particular lung cancer, comprising administering to an individual in need thereof an effective amount of an agent which decreases the expression of an informative gene whose expression is increased in said individual as compared with a normal individual.

25 The invention also relates to a method of treating a disease of the lung, in particular lung cancer, comprising administering to an individual in need thereof an effective amount of an agent which increases the activity of an expression product of an informative gene whose activity is decreased in said individual as compared with a normal individual.

30 The invention also relates to a method of treating a disease of the lung, in particular lung cancer, comprising administering to an individual in need thereof an

-6-

effective amount of an agent which decreases the activity of an expression product of an informative gene whose activity is increased in said individual as compared with a normal individual.

The invention also provides a substrate for use in expression analyses having  
5 immobilized thereon a plurality of polymers (e.g., oligonucleotides, polypeptides, etc.) which bind specifically to one or more informative genes disclosed herein or expression products of said genes. In a particular embodiment, the substrate is an array (e.g., a microarray) having immobilized thereon a plurality of oligonucleotides. In one embodiment, less than 1000 oligonucleotides or polymers are immobilized on  
10 the array. In another embodiment, the oligonucleotides hybridize specifically to one allelic form of one or more informative genes disclosed herein. In a particular embodiment, the informative genes are selected from the group consisting essentially of the genes shown in FIGS. 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E. In one embodiment the substrate has immobilized thereon  
15 only polymers which bind specifically to one or more of the informative genes disclosed herein or expression products of said genes. In other embodiments the substrate has additional polymers immobilized thereon, such as polymers which bind specifically to informative genes for disorders such as types of cancer.

#### BRIEF DESCRIPTION OF THE DRAWINGS

20 The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings.

FIGS. 1A-1N is a table of genes which are differentially expressed in smokers and smokers with cancer. Column one is the probe set designation for the  
25 gene; Column two is the common name(s) for the gene; Column three is the GenBank accession number for the gene sequence; Column four is the map (chromosomal) position of the gene; and Column five is descriptive information for the gene. The genes which overlap between FIGS. 1A-1N and FIGS. 2A-2F are indicated with an “^” (21 genes). The genes which overlap between FIGS. 1A-1N  
30 and FIGS. 3A-3D are indicated with an “\*” (5 genes).

FIGS. 2A-2F is a table of genes which are differentially expressed in smokers with cancer and non- smokers. Column one is the probe set designation for the gene; Column two is the common name(s) for the gene; Column three is the GenBank accession number for the gene sequence; Column four is the map (chromosomal) position of the gene; and Column five is descriptive information for the gene. The genes which overlap between FIGS. 1A-1N and FIGS. 2A-2F are indicated with an “^” (21 genes). The genes which overlap between FIGS. 2A-2F and FIGS. 3A-3D are indicated with an “#” (23 genes).

FIGS. 3A-3D is a table of genes which are differentially expressed in smokers and non- smokers. Column one is the probe set designation for the gene; Column two is the common name(s) for the gene; Column three is the GenBank accession number for the gene sequence; Column four is the map (chromosomal) position of the gene; and Column five is descriptive information for the gene. The genes which overlap between FIGS. 3A-3D and FIGS. 1A-1N are indicated with an “\*” (5 genes). The genes which overlap between FIGS. 3A-3D and FIGS. 2A-2F are indicated with an “#” (23 genes).

FIGS. 4A-4P is a table of 131 genes which are differentially expressed and can be used to distinguish between non-smokers, smokers and smokers with cancer. Column one is the fold change in expression between non-smokers and smokers; Column two shows how the expression differs in smokers as compared to non-smokers; Column three is the fold change in expression between non-smokers and smokers with cancer; Column four shows how the expression differs in smokers with cancer as compared to non-smokers; Column five is the fold change in expression between smokers and smokers with cancer; Column six shows how the expression differs in smokers with cancer as compared to smokers; Column seven is the common name(s) for the gene; Column eight is the GenBank accession number for the gene sequence; Column nine is the map position of the gene; Column ten is the chromosomal position of the gene; Column 11 is the description of the gene.

FIGS. 5A-5F is a table of data similar to the data shown in FIGS. 4A-4P for 131 genes which are differentially expressed and can be used to distinguish between non-smokers, smokers and smokers with cancer. Column one is the probe set

designation for the gene; Column two is the common name(s) for the gene; Column three is the GenBank accession number for the gene sequence; Column four is the map (chromosomal) position of the gene; and Column five is descriptive information for the gene.

5        FIGS. 6A-6R is a table of genes which are differentially expressed in smokers and non-smokers. Column one is the probe set designation for the gene; Column two is the common name(s) for the gene; Column three is the GenBank accession number for the gene sequence; Column four is the p value for the smoker/non-smoker comparison; Column five shows whether the gene is up or down  
10 regulated; Column six shows the fold change in expression; Column seven is the GenBank description of the gene; Column eight is the functional category of the gene as indicated in GenBank; Column nine is the map (chromosomal) position of the gene; Column ten is the link locus; Column eleven is the Unigene reference for the gene; and Column twelve is OMIM reference for the gene. On FIG. 6R, five  
15 probe sets are shown which resulted in redundant results.

FIGS. 7A-7L is a table of genes which are differentially expressed in samples from smokers (S), nonsmokers (NS) and smokers with cancer (C). Column one is the probe set designation for the gene; Column two is the gene title; Column three is the gene symbol; Column four is the chromosomal location of the gene; Column five  
20 is the locus link.

FIGS. 8A-8E is a table of genes which are differentially expressed in samples from smokers (S) and smokers with cancer (C). Column one is the probe set designation for the gene; Column two is the gene title; Column three is the gene symbol; Column four is the chromosomal location of the gene; Column five is the  
25 locus link.

## DETAILED DESCRIPTION OF THE INVENTION

A description of preferred embodiments of the invention follows.

Lung cancer in smokers involves histopathological and molecular progression from normal to premalignant to cancer. Gene expression arrays of lung  
30 tumors have been used to characterize expression profiles of lung cancers, and to

show the progression of molecular changes from non-malignant lung tissue to lung cancer. The ability to determine which individuals, particularly among smokers, have molecular changes in their airway epithelial cells and how these changes relate to premalignant and malignant changes that have been described in the lung has  
5 great potential for determining risk and for diagnosing cancer at a stage when treatment can be more effective, thus reducing the mortality and morbidity rates of lung cancer. The relative ease with which airway epithelial cells can be obtained (from bronchoscopy and buccal mucosal scrapings) suggests that this approach will have wide clinical applicability and has the potential for becoming a standard  
10 clinical screening tool for the large number of subjects at risk for developing disorders of the lung.

As used in this application, the singular form "a," "an," and "the" include plural references unless the context clearly dictates otherwise. For example, the term "an agent" includes a plurality of agents, including mixtures thereof.

15 An individual is not limited to a human being but may also be other organisms including but not limited to mammals, plants, bacteria, or cells derived from any of the above.

Throughout this disclosure, various aspects of this invention can be presented in a range format. It should be understood that the description in range format is  
20 merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed  
25 subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

The practice of the present invention may employ, unless otherwise indicated, conventional techniques and descriptions of organic chemistry, polymer  
30 technology, molecular biology (including recombinant techniques), cell biology, biochemistry, and immunology, which are within the skill of the art. Such

-10-

conventional techniques include polymer array synthesis, hybridization, ligation, and detection of hybridization using a label. Specific illustrations of suitable techniques can be had by reference to the example herein below. However, other equivalent conventional procedures can, of course, also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals such as *Genome Analysis: A Laboratory Manual Series (Vols. I-IV)*, *Using Antibodies: A Laboratory Manual*, *Cells: A Laboratory Manual*, *PCR Primer: A Laboratory Manual*, and *Molecular Cloning: A Laboratory Manual* (all from Cold Spring Harbor Laboratory Press), Stryer, L. (1995) *Biochemistry* (4th Ed.) Freeman, New York, Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London, Nelson and Cox (2000), *Lehninger, Principles of Biochemistry* 3<sup>rd</sup> Ed., W.H. Freeman Pub., New York, NY and Berg *et al.* (2002) *Biochemistry*, 5<sup>th</sup> Ed., W.H. Freeman Pub., New York, NY, all of which are herein incorporated in their entirety by reference for all purposes.

The present invention can employ solid substrates, including arrays in some preferred embodiments. Methods and techniques applicable to polymer (including protein) array synthesis have been described in U.S.S.N 09/536,841, WO 00/58516, U.S. Patents Nos. 5,143,854, 5,242,974, 5,252,743, 5,324,633, 5,384,261, 5,405,783, 5,424,186, 5,451,683, 5,482,867, 5,491,074, 5,527,681, 5,550,215, 5,571,639, 5,578,832, 5,593,839, 5,599,695, 5,624,711, 5,631,734, 5,795,716, 5,831,070, 5,837,832, 5,856,101, 5,858,659, 5,936,324, 5,968,740, 5,974,164, 5,981,185, 5,981,956, 6,025,601, 6,033,860, 6,040,193, 6,090,555, 6,136,269, 6,269,846 and 6,428,752, in PCT Applications Nos. PCT/US99/00730 (International Publication Number WO 99/36760) and PCT/US01/04285 (International Publication Number WO 01/58593), which are all incorporated herein by reference in their entirety for all purposes.

The term "array" as used herein refers to an intentionally created collection of molecules which can be prepared either synthetically or biosynthetically. The molecules in the array can be identical or different from each other. The array can assume a variety of formats, for example, libraries of soluble molecules; libraries of compounds tethered to resin beads, silica chips, or other solid supports.

-11-

Additionally, the term "array" is meant to include those libraries of nucleic acids which can be prepared by spotting nucleic acids of essentially any length (for example, from 1 to about 1000 nucleotide monomers in length) onto a substrate.

The term "solid support", "support", and "substrate" as used herein are used interchangeably and refer to a material or group of materials having a rigid or semi-rigid surface or surfaces. In many embodiments, at least one surface of the solid support will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different compounds with, for example, wells, raised regions, pins, etched trenches, or the like. According to other embodiments, the solid support(s) will take the form of beads, resins, gels, microspheres, or other geometric configurations. See U.S. Patent No. 5,744,305 for exemplary substrates.

Patents that describe synthesis techniques in specific embodiments include U.S. Patents Nos. 5,412,087, 6,147,205, 6,262,216, 6,310,189, 5,889,165, and 5,959,098. Nucleic acid arrays are described in many of the above patents, but the same techniques are applied to polypeptide arrays.

Nucleic acid arrays that are useful in the present invention include those that are commercially available from Affymetrix (Santa Clara, CA) under the brand name GeneChip®.

The present invention also contemplates many uses for polymers attached to solid substrates. These uses include gene expression monitoring, profiling, library screening, genotyping and diagnostics. Gene expression monitoring, and profiling methods can be shown in U.S. Patents Nos. 5,800,992, 6,013,449, 6,020,135, 6,033,860, 6,040,138, 6,177,248 and 6,309,822. Genotyping and uses therefore are shown in U.S.S.N. 60/319,253, 10/013,598 (U.S. Patent Application Publication No.: 20030036069), and U.S. Patents Nos. 5,856,092, 6,300,063, 5,858,659, 6,284,460, 6,361,947, 6,368,799 and 6,333,179. Other uses are embodied in U.S. Patents Nos. 5,871,928, 5,902,723, 6,045,996, 5,541,061, and 6,197,506.

The present invention also contemplates sample preparation methods in certain preferred embodiments. Prior to or concurrent with genotyping, the genomic sample may be amplified by a variety of mechanisms, some of which may employ



-12-

PCR. See, e.g., *PCR Technology: Principles and Applications for DNA Amplification* (Ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (Eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (Eds. McPherson et al., IRL Press, Oxford); and U.S. Patent Nos. 4,683,202, 4,683,195, 4,800,159, 4,965,188, and 5,333,675, and each of which is incorporated herein by reference in their entireties for all purposes. The sample may be amplified on the array. See, for example, U.S. Patent No. 6,300,070 and U.S. Patent Application No. 09/513,300, which are  
10 incorporated herein by reference.

Other suitable amplification methods include the ligase chain reaction (LCR) (e.g., Wu and Wallace, *Genomics* 4, 560 (1989), Landegren et al., *Science* 241, 1077 (1988) and Barringer et al. *Gene* 89:117 (1990)), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989) and WO88/10315), self-sustained  
15 sequence replication (Guatelli et al., *Proc. Nat. Acad. Sci. USA*, 87, 1874 (1990) and WO90/06995), selective amplification of target polynucleotide sequences (U.S. Patent No. 6,410,276), consensus sequence primed polymerase chain reaction (CP-PCR) (U.S. Patent No 4,437,975), arbitrarily primed polymerase chain reaction (AP-PCR) (U.S. Patent No 5,413,909, 5,861,245) and nucleic acid based sequence  
20 amplification (NABSA). (See, U.S. Patents Nos. 5,409,818, 5,554,517, and 6,063,603, each of which is incorporated herein by reference). Other amplification methods that may be used are described in, U.S. Patent Nos. 5,242,794, 5,494,810, 4,988,617 and in U.S.S.N. 09/854,317, each of which is incorporated herein by reference.

25 Additional methods of sample preparation and techniques for reducing the complexity of a nucleic sample are described in Dong et al., *Genome Research* 11, 1418 (2001), in U.S. Patent Nos. 6,361,947, 6,391,592 and U.S. Patent Application Nos. 09/916,135, 09/920,491 (U.S. Patent Application Publication No. 20030096235), 09/910,292 (U.S. Patent Application Publication No. 20030082543,  
30 U.S. Patent No. 6,632,611), and 10/013,598 (U.S. Patent Application Publication No. 20030036069).

Methods for conducting polynucleotide hybridization assays have been well developed in the art. Hybridization assay procedures and conditions will vary depending on the application and are selected in accordance with the general binding methods known including those referred to in: Maniatis *et al. Molecular Cloning: A Laboratory Manual* (2<sup>nd</sup> Ed. Cold Spring Harbor, N.Y, 1989); Berger and Kimmel *Methods in Enzymology*, Vol. 152, *Guide to Molecular Cloning Techniques* (Academic Press, Inc., San Diego, CA, 1987); Young and Davism, *P.N.A.S.*, 80: 1194 (1983). Methods and apparatus for carrying out repeated and controlled hybridization reactions have been described in U.S. Patent Nos. 5,871,928, 5,874,219, 6,045,996 and 6,386,749, 6,391,623, each of which are incorporated herein by reference.

The present invention also contemplates signal detection of hybridization between ligands in certain preferred embodiments. See U.S. Patent Nos. 5,143,854, 5,578,832; 5,631,734; 5,834,758; 5,936,324; 5,981,956; 6,025,601; 6,141,096; 6,185,030; 6,201,639; 6,218,803; and 6,225,625, in U.S. Patent Application No. 60/364,731 and in PCT Application PCT/US99/06097 (published as WO99/47964), each of which also is hereby incorporated by reference in its entirety for all purposes.

Methods and apparatus for signal detection and processing of intensity data are disclosed in, for example, U.S. Patents Nos. 5,143,854, 5,547,839, 5,578,832, 5,631,734, 5,800,992, 5,834,758; 5,856,092, 5,902,723, 5,936,324, 5,981,956, 6,025,601, 6,090,555, 6,141,096, 6,185,030, 6,201,639; 6,218,803; and 6,225,625, in U.S. Patent Application No. 60/364,731 and in PCT Application PCT/US99/06097 (published as WO99/47964), each of which also is hereby incorporated by reference in its entirety for all purposes.

The practice of the present invention may also employ conventional biology methods, software and systems. Computer software products of the invention typically include computer readable medium having computer-executable instructions for performing the logic steps of the method of the invention. Suitable computer readable medium include floppy disk, CD-ROM/DVD/DVD-ROM, hard-disk drive, flash memory, ROM/RAM, magnetic tapes and etc. The computer executable instructions may be written in a suitable computer language or

combination of several languages. Basic computational biology methods are described in, e.g. Setubal and Meidanis *et al.*, *Introduction to Computational Biology Methods* (PWS Publishing Company, Boston, 1997); Salzberg, Searles, Kasif, (Ed.), *Computational Methods in Molecular Biology*, (Elsevier, Amsterdam, 5 1998); Rashidi and Buehler, *Bioinformatics Basics: Application in Biological Science and Medicine* (CRC Press, London, 2000) and Ouelette and Bzevanis *Bioinformatics: A Practical Guide for Analysis of Gene and Proteins* (Wiley & Sons, Inc., 2<sup>nd</sup> ed., 2001).

The present invention may also make use of various computer program 10 products and software for a variety of purposes, such as probe design, management of data, analysis, and instrument operation. See, U.S. Patent Nos. 5,593,839, 5,795,716, 5,733,729, 5,974,164, 6,066,454, 6,090,555, 6,185,561, 6,188,783, 6,223,127, 6,229,911 and 6,308,170.

Additionally, the present invention may have preferred embodiments that 15 include methods for providing genetic information over networks such as the Internet as shown in U.S. Patent Application Nos. 10/063,559 (U.S. Patent Application Publication No. 20020183936), 60/349,546, 60/376,003, 60/394,574, 60/403,381.

Further methods for practicing the present invention are also exemplified in 20 Spira *et al.*, *Chest* 125(5 Suppl):115S (2004) ; Spira *et al.*, *Biotechniques* 36(3):484-7 (2004) ; and Powell *et al.*, *Am. J. Respir. Cell. Mol. Biol.* 29(2):157-62 (2003), the teachings of all of which are incorporated herein in their entirety.

As used herein, an informative gene is a gene which is differentially expressed in sample populations with different phenotypes (e.g., presence of or 25 susceptibility to disorders of the lung, particularly lung cancer) or behavioral traits (e.g., smoking), e.g., differentially expressed in smokers as compared with non-smokers, in smokers with cancer as compared to smokers without cancer, in smokers likely to develop cancer as compared with smokers unlikely to develop cancer, etc. In one embodiment, an informative gene displays at least a fold change shown in one 30 of the rows of Column six of FIGS. 6A-6R. In a preferred embodiment, an informative gene displays at least a two-fold, more preferably at least a five-fold, and

even more preferably at least a ten-fold difference in expression between two sample populations. Examples of informative genes are shown in FIGS. 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E; other informative genes can be identified using methods described herein.

5       As used herein, gene expression products are proteins, peptides, or nucleic acid molecules (e.g., mRNA, tRNA, rRNA, or cRNA) that are involved in transcription or translation. The nucleic acid molecule levels measured can be derived directly from the gene or, alternatively, from a corresponding regulatory gene. All forms of gene expression products can be measured, including, for  
10       example, spliced variants. Similarly, gene expression can be measured by assessing the level of protein or derivative thereof translated from mRNA. In this embodiment the determination of the gene expression profile can be made using techniques for protein detection and quantitation known in the art. For example, antibodies specific for the protein or polypeptide can be obtained using methods which are routine in  
15       the art, and the specific binding of such antibodies to protein or polypeptide gene expression products can be detected and measured. Alternatively, polymers immobilized on a substrate (e.g., protein arrays) can be used.

      In the embodiment wherein arrays of nucleic acids are immobilized on a surface, the number of nucleic acid sequences may be selected for different  
20       applications, and may be, for example, about 100 or more, or, e.g., in some embodiments, more than  $10^5$  or  $10^8$ . In one embodiment, the surface comprises at least 100 probe nucleic acids each preferably having a different sequence, each probe contained in an area of less than about  $0.1 \text{ cm}^2$ , or, for example, between about  $1 \text{ um}^2$  and  $10,000 \text{ um}^2$ , and each probe nucleic acid having a defined sequence and  
25       location on the surface. In one embodiment, at least 1,000 different nucleic acids are provided on the surface, wherein each nucleic acid is contained within an area less than about  $10^{-3} \text{ cm}^2$ , as described, for example, in U.S. Patent No. 5,510,270, the disclosure of which is incorporated herein.

      Arrays of nucleic acids for use in gene expression monitoring are described  
30       in PCT WO 97/10365, the disclosure of which is incorporated herein. In one embodiment, arrays of nucleic acid probes are immobilized on a surface, wherein the

array comprises more than 100 different nucleic acids and wherein each different nucleic acid is localized in a predetermined area of the surface, and the density of the different nucleic acids is greater than about 60 different nucleic acids per  $1 \text{ cm}^2$ .

Arrays of nucleic acids immobilized on a surface which may be used also are described in detail in U.S. Patent No. 5,744,305, the disclosure of which is incorporated herein. As disclosed therein, on a substrate, nucleic acids with different sequences are immobilized each in a predefined area on a surface. For example, 10, 50, 60, 100,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ , or  $10^8$  different monomer sequences may be provided on the substrate. The nucleic acids of a particular sequence are provided within a predefined region of a substrate, having a surface area, for example, of about  $1 \text{ cm}^2$  to  $10^{-10} \text{ cm}^2$ . In some embodiments, the regions have areas of less than about  $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$ ,  $10^{-5}$ ,  $10^{-6}$ ,  $10^{-7}$ ,  $10^{-8}$ ,  $10^{-9}$ , or  $10^{-10} \text{ cm}^2$ . For example, in one embodiment, there is provided a planar, non-porous support having at least a first surface, and a plurality of different nucleic acids attached to the first surface at a density exceeding about 400 different nucleic acids/ $\text{cm}^2$ , wherein each of the different nucleic acids is attached to the surface of the solid support in a different predefined region, has a different determinable sequence, and is, for example, at least 4 nucleotides in length. The nucleic acids may be, for example, about 4 to 20 nucleotides in length. The number of different nucleic acids may be, for example, 1000 or more. In the embodiment where polynucleotides of a known chemical sequence are synthesized at known locations on a substrate, and binding of a complementary nucleotide is detected, and wherein a fluorescent label is detected, detection may be implemented by directing light to relatively small and precisely known locations on the substrate. For example, the substrate is placed in a microscope detection apparatus for identification of locations where binding takes place. The microscope detection apparatus includes a monochromatic or polychromatic light source for directing light at the substrate, means for detecting fluoresced light from the substrate, and means for determining a location of the fluoresced light. The means for detecting light fluoresced on the substrate may in some embodiments include a photon counter. The means for determining a location of the fluoresced light may include an x/y translation table for the substrate.

Translation of the substrate and data collection are recorded and managed by an appropriately programmed digital computer, as described in U.S. Patent No. 5,510,270, the disclosure of which is incorporated herein.

The sample to be assessed can be any sample that contains a gene expression  
5 product. Suitable sources of gene expression products, i.e., samples, can include cells, lysed cells, cellular material for determining gene expression, or material containing gene expression products. Examples of such samples are blood, plasma, lymph, urine, tissue, mucus, sputum, saliva or other cell samples. Methods of  
10 obtaining such samples are known in the art. In a preferred embodiment, the sample is an airway epithelial cell.

Disorders which may be diagnosed or treated by methods described herein include, but are not limited to, asthma, chronic bronchitis, emphysema, bronchiectasis, primary pulmonary hypertension and acute respiratory distress  
15 syndrome. The methods described herein may also be used to diagnose or treat lung disorders that involve the immune system including, hypersensitivity pneumonitis, eosinophilic pneumonias and persistent fungal infections. Other disorders include pulmonary fibrosis, systemic sclerosis, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, cancers of the lung such as adenocarcinoma, squamous cell carcinoma, small cell and large cell carcinomas, and benign  
20 neoplasms of the lung including bronchial adenomas and hamartomas.

The present invention also provides methods for monitoring the effect of a treatment regimen in an individual by monitoring the expression for one or more informative genes. For example, a baseline expression level for one or more informative genes from the individual can be determined, and repeated gene  
25 expression levels can be determined at time points during treatment. A shift in gene expression from a profile correlated with the disorder or with poor treatment outcome to a profile correlated with lack of the disorder or improved treatment outcome is evidence of an effective therapeutic regimen, while a repeated profile correlated with the disorder or with poor treatment outcome is evidence of an  
30 ineffective therapeutic regimen.

It is also clear that the present invention can be used to generate databases comprising informative genes which will have many applications in medicine, research and industry and which can be used as a point of comparison in many of the analyses described herein.

5       The present invention has many preferred embodiments and relies on many patents, applications and other references for details known to those of the art. Therefore, when a patent, application, or other reference is cited herein, it should be understood that it is incorporated by reference in its entirety for all purposes as well as for the proposition that is recited.

## 10   EXAMPLES

### *Example 1*

#### **Methods**

Samples of epithelial cells, obtained by brushing airway surfaces, were obtained from intra- and extra-pulmonary airways in 11 normal non-smokers (NS),  
15 15 smokers without lung cancer (S), and 9 smokers with lung cancer (SC). 5-10 µg of RNA was extracted using standard trizol-based methods, quality of RNA was assayed in gels, and the RNA was processed using standard protocols developed by Affymetrix for the U133 human array. Expression profiles, predictive algorithms, and identification of critical genes are made using bioinformatic methods. In  
20 particular, "Significance analysis of microarrays" (SAM) software (Tusher *et al.*, *PNAS*, 98:5116-21, 2001) and "Prediction analysis for microarrays" (PAM) software (Tibshirani *et al.*, *PNAS*, 99:6567-72, 2002) (both Stanford University Proteomics Center) were used.

For the analysis shown in FIGS. 4A-4P, samples were obtained as described  
25 above from 6 smokers with lung cancer (SC), 9 smokers without lung cancer (S) and 5 normal non-smokers (NS).

## Results

FIGS. 1A-1N show genes which are differentially expressed in smokers and smokers with cancer; FIGS. 2A-2F show genes which are differentially expressed in smokers with cancer and non- smokers; and FIGS. 3A-3D show genes which are  
5 differentially expressed in smokers and non- smokers.

FIGS. 4A-4P and 5A-5F show 131 genes which are differentially expressed and can be used to distinguish between non-smokers, smokers and smokers with cancer.

FIGS. 6A-6R show genes which are differentially expressed in smokers and  
10 non-smokers.

FIGS. 7A-7L is a table of genes which are differentially expressed in samples from smokers (S), nonsmokers (NS) and smokers with cancer (C).

FIGS. 8A-8E is a table of genes which are differentially expressed in samples from smokers (S) and smokers with cancer (C).

15

The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in  
20 the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.



-20-

## CLAIMS

What is claimed is:

1. A method of diagnosing a lung cancer comprising:
  - a) obtaining a nucleic acid sample from an individual to be diagnosed;  
5 and
  - b) determining the expression of one or more informative genes in said sample, wherein the informative genes are selected from the group consisting of the genes shown in Figures 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E,  
10 wherein increased expression of an informative gene whose expression is increased in individuals having a lung cancer, or decreased expression of an informative gene whose expression is decreased in individuals having a lung cancer, is indicative of a lung cancer in the individual.
2. A method according to Claim 1, wherein the lung cancer selected from the  
15 group consisting of adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, and benign neoplasms of the lung.
3. A method according to Claim 2, wherein the lung cancer is a benign neoplasm selected from the group consisting of bronchial adenomas and hamartomas.
- 20 4. A method according to Claim 1, wherein the nucleic acid sample is RNA.
5. A method according to Claim 1, wherein the individual to be diagnosed is an individual who has been exposed to tobacco smoke.
6. A method according to Claim 1, wherein the individual to be diagnosed is an individual who has smoked.

-21-

7. A method according to Claim 1, wherein the individual to be diagnosed is an individual who smokes.
8. A method according to Claim 1, wherein the expression of two or more informative genes is determined.
- 5 9. A method according to Claim 1, wherein the expression of five or more informative genes is determined.
10. A method according to Claim 1, wherein the expression of ten or more informative genes is determined.
11. A method according to Claim 1, wherein the expression of fifteen or more  
10 informative genes is determined.
12. A method according to Claim 1, wherein the expression of twenty or more informative genes is determined.
13. A method according to Claim 1, wherein the expression of fifty or more informative genes is determined.
- 15 14. A method according to Claim 1, wherein the expression of one hundred or more informative genes is determined.
15. A method according to Claim 1, wherein the nucleic acid sample is obtained from an airway epithelial cell.
16. A method according to Claim 15, wherein the airway epithelial cell is  
20 obtained from a bronchoscopy or buccal mucosal scraping.

-22-

17. A method according to Claim 1, wherein the expression is determined using a microarray having one or more probes for said one or more genes immobilized thereon.
18. A method of treating a lung cancer comprising administering to an individual  
5 in need thereof an effective amount of an agent which increases the expression of an informative gene whose expression is decreased in said individual as compared with a normal individual, wherein the informative gene is selected from the group consisting of the genes shown in Figures 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E.
- 10 19. A method of treating a lung cancer comprising administering to an individual in need thereof an effective amount of an agent which decreases the expression of an informative gene whose expression is increased in said individual as compared with a normal individual, wherein the informative  
15 gene is selected from the group consisting of the genes shown in Figures 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E.
20. A method of treating a lung cancer comprising administering to an individual in need thereof an effective amount of an agent which increases the activity of an expression product of an informative gene whose activity is decreased in said individual as compared with a normal individual, wherein the  
20 informative gene is selected from the group consisting of the genes shown in Figures 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E.
21. A method of treating a lung cancer comprising administering to an individual in need thereof an effective amount of an agent which decreases the activity of an expression product of an informative gene whose activity is increased  
25 in said individual as compared with a normal individual, wherein the informative gene is selected from the group consisting of the genes shown in Figures 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E.

22. A substrate having immobilized thereon a plurality of less than one thousand polymers which bind specifically to an expression product of one or more genes selected from the group consisting essentially of the genes shown in Figures 1A-1N, 2A-2F, 3A-3D, 4A-4P, 5A-5F, 6A-6R, 7A-7L and 8A-8E.

53003_analysis s							
Systematic	Common	Genbank	Map	Description			
206916_x_at	TAT	NM_000353	16q22.1	tyrosine aminotransferase (TAT), nuclear gene encoding mitochondrial protein			
212217_at	KIAA0436	AB007896	2p22.1	putative L-type neutral amino acid transporter			
212238_at	KIAA0978	AL117518	20q11.1	hypothetical protein KIAA0978 protein			
202230_s_at	CHERP	NM_006387	19p13.1	polyglutamine repeat; calcium (ca2+) homeostasis endoplasmic reticulum protein (ERROT213-21)			
202833_s_at	SERPINA1	NM_000295	14q32.1	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1 (SERPINA1)			
211686_s_at	LOC84549	AF251062	8p11.23	RNA binding protein			
211800_s_at	USP4	AF017306	3p21.3	deubiquitinating enzyme UnpES (UNP) ubiquitin specific protease 4 (proto-oncogene)			
213429_at		AW025579	10	ESTs			
213515_x_at	HBG1	AI133353	11p15.5	hemoglobin, gamma A			
^213851_at		BG031677	3	ESTs, Weakly similar to S65173 probable membrane protein YPL162c - yeast			
212639_x_at	K-ALPHA-1	AL581768	12q12-12q14.3	tubulin, alpha, ubiquitous			
213196_at	KIAA0326	AI924293	16p11.1	KIAA0326 protein			
213266_at	76P	BF592982	15q15	gamma tubulin ring complex protein (76p gene)			
214214_s_at	C1QBP	AU151801	17p13.3	complement component 1, q subcomponent binding protein			
^208514_at	KCNE1	NM_000219	21q22.12	potassium voltage-gated channel, Isk-related family			
208922_s_at	INXF1	BC004904	11q12-q13	nuclear RNA export factor 1			
203953_s_at	CLDN3	BE791251	7q11.23	claudin 3			

FIG. 1A

206861_s_at	CGGBP1	NM_003663	3p12-p11.1	CGG triplet repeat binding protein 1 (CGGBP1)
205130_at	RAGE	NM_014226	14q32	renal tumor antigen (RAGE)
	DKFZP586D22			
206961_s_at	23	NM_004275	6pter-p12.1	TRF-proximal protein (TRFP)
210513_s_at	VEGF	AF091352	6p12	vascular permeability factor 148
211251_x_at	NFYC	U78774	1p32	NFY-C nuclear transcription factor Y, gamma
*203136_at	RABAC1	NM_006423	19q13.31	Rab acceptor 1 (prenylated) (RABAC1)
209263_x_at	TM4SF7	BC000389	11	transmembrane 4 superfamily member 7
210438_x_at	SSA2	M25077	1q31	SS-ARo ribonucleoprotein autoantigen 60 kd subunit
203579_s_at	FLJ13291	NM_003983	16q22.2	solute carrier family 7 (cationic amino acid transporter, y+ system), member 6
219327_s_at	GPRC5C	NM_022036	17q25	G protein-coupled receptor, family C, group 5, member C (GPRC5C), transcript variant 1
219664_s_at	DECR2	NM_020664	16p13.12	2,4-dienoyl CoA reductase 2, peroxisomal (DECR2)
220045_at	NEUROD6	NM_022728	7p15.3	neurogenic differentiation 6 (NEUROD6)
*218321_x_at	LOC51657	NM_016086	7p12.3	map kinase phosphatase-like protein MK-STYX (LOC51657)
218574_s_at	LMCD1	NM_014583	3p26-p24	LIM and cysteine-rich domains 1 (LMCD1)
*218766_s_at	WARS2	NM_015836	1p13.3-p13.1	tryptophanyl tRNA synthetase 2 (mitochondrial) (WARS2), nuclear gene encoding mitochondrial protein
221037_s_at	DKFZP434N12			
	35	NM_031291	4q26	hypothetical protein DKFZp434N1235
221539_at	EIF4EBP1	AB044548	8p12	POKcl.6 mRNA for eukaryotic translation initiation factor 4E binding protein 1
49051_g_at		N92708		BRACE2015132, weakly similar to Drosophila melanogaster Oregon R cytoplasmic basic protein (deltex)

FIG. 1B

201087 at	PXN	NM_002859	12q24	paxillin (PXN).
220744 s at	WDR10	NM_018262	3q21	hypothetical protein FLJ10897 (FLJ10897)
200976 s at	NUDT1	NM_006024	7p22	Tax1 (human T-cell leukemia virus type I) binding protein 1 (TAX1BP1)
59705 at	SCLY	AA911739	2q37.3	putative selenocysteine lyase
215984 s at		AL121845		KIAA1088 gene
215357 s at		Z93241		chromosome 22 Contains three novel genes, a Mitochondrial ATP Synthase G Chain (EC 3.6.1.34) pseudogene and the DIA1 gene for diaphorase (NADH) (cytochrome b-5 reductase)
202943 s at	NAGA	M38083	22q13-qter	alpha-N-acetylgalactosaminidase
216502 at	FLJ12671	AL096734	1q21.2	Homo sapiens, Similar to hypothetical protein FLJ12484
216475 at	dJ468K3.1	AL133269	q25-26	sequence from clone RP3-468K3 on chromosome 6q25-26. Contains a beta tubulin pseudogene
207508 at	ATP5G3	NM_001689	2q31.1	ATP synthase, H+ transporting, mitochondrial F0 complex
207196 s at	TNIP1	NM_006058	5q32-q33.1	Nef-associated factor 1 (NAF1)
214843 s at	VDU1	AK022864	1p31.1	weakly similar to UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 4
208610 s at	SRRM2	AI655799	16p13.3	serine/arginine repetitive matrix 2
221518 s at	FLJ20727	BC000226	11p15.2	hypothetical protein FLJ20727 /FL=gb:BC000226.1
212027 at		BE466128	14	Homo sapiens cDNA: FLJ22454 fis, clone HRC09703
206555 s at	FLJ20274	NM_017736	16p12.2	protein FLJ20274 (FLJ20274)
206055 s at	SNRPA1	NM_003090	15q26.3	small nuclear ribonucleoprotein polypeptide A (SNRPA1)

FIG. 1C

214728 x at	SMARCA4; BRG1; BAF190; SNF2L4; SNF2LB; hSNF2B; hSNF2b; SNF2-				
221651 x at	BETA	AK026573	19p13.2		highly similar to HSU29175 Human transcriptional activator (BRG1)
214709 s at	IGKC	BC005332	2p12		Similar to immunoglobulin kappa constant
	KTN1	Z22551	14q22.1		kinectin gene
^219156 at	SYNJ2BP	NM_018373	14q22.1-q24.3		hypothetical protein FLJ11271
212341 at	GTF2A2	AA195936	15q21.2		general transcription factor IIA, 2, 12kDa
215599 at	SMA3	X83300	5q13		SMA4
211503 s at	RAB14	AF112206	9q32-q34.11		ras-related protein rab-14
211430 s at		M87789			anti-hepatitis A IgG variable region
209422 at		AL109965			the hepatocellular carcinoma-associated antigen 58 (HCA58) and a putative novel protein containing a PHD finger domain)
208933 s at	FCGR2B	AI659005	1q23		Fc fragment of IgG, low affinity IIb, receptor for (CD32)
214989 x at	PEPP2	BC000969	12p12		phosphoinositol 3-phosphate-binding protein-2
219022 at	FLJ12448	NM_022895	12q		hypothetical protein FLJ12448
^212977 at	RDC1	AI817041	2q37.3		G protein-coupled receptor
209682 at	CBLB	U26710	3q13.12		cbl-b
202378 s at	HSOBRGRP	NM_017526	1		leptin receptor gene-related protein
202082 s at	SEC14L1	NM_003003	17q25.1-17q25.2		
202073 at	OPTN	AV757675	10p12.33		optineurin

FIG. 1D



203595 s at	RI58	N47725	10q23.33	retinoic acid- and interferon-inducible protein (58kD)
201019 s at	EIF1A	NM_001412	X	translation initiation factor 1A (EIF1A)
201222 s at	RAD23B	AL527365	9q31.2	RAD23 homolog B (S. cerevisiae)
*201833 at	HDAC2	NM_001527	6q21	histone deacetylase 2 (HDAC2)
201488 x at	KHDRBS1	BC000717	1p32	GAP-associated tyrosine phosphoprotein p62 (Sam68)
202880 s at	PSCD1	NM_004762	17q25	pleckstrin homology, Sec7 and coiledcoil
205718 at	ITGB7	NM_000889	12q13.13	integrin, beta 7 (ITGB7)
201901 s at	YY1	NM_003403	14q	YY1 transcription factor
201991 s at	KIF5B	BF223224	10pter-q22.1	kinesin family member 5B
210259 s at	DLX4	AF254115	17q21.33	beta protein 1 BP1
220237 at	APG3	NM_022488	3q13.31	PC3-96 protein (PC3-96)
209700 x at	PDE4DIP	AB042555	1q12	similar to rat myomegalin
203437 at	PMI	NM_003876	17p13.1	putative receptor protein (PMI)
^214685 at		AB000464	4p16.3	clone:RES4-24A
220949 s at	FLJ11000	NM_024033	7q31.32	hypothetical protein MGC5242 (MGC5242)
213528 at	MGC9084	AL035369	1q22	novel gene from PAC 117P20, chromosome 1
208986 at	TCF12	AL559478	15q21	transcription factor 12 (HTF4, helix-loop-helix
221145 at		NM_018499		transcription factors 4)
205982 x at	SFTPC	NM_003018	8p21	hypothetical protein PRO1097
210869 s at	MCAM	M29277	11q23.3	surfactant, pulmonary-associated protein C (SFTPC)
218275 at	SLC25A10	NM_012140	17q25.3	MUC18 glycoprotein melanoma adhesion molecule /FL=gb:M29277.1
207971 s at	RAB1A	NM_015147	2p14	solute carrier family 25 (mitochondrial carrier, dicarboxylate transporter)
208345 s at	POU3F1	NM_002699	1p34.1	KIAA0582 protein KIAA0582
				POU domain, class 3, transcription factor 1 (POU3F1)

FIG. 1E

^216072 at		AL050065	21q	DKFZp566M043
202419 at	FVT1	NM_002035	18q21.3	follicular lymphoma variant translocation 1 (FVT1)
217118 s at	KIAA0930	AK025608	22q13.31	Consensus includes gb:AK025608.1 /DEF=Homo sapiens cDNA: FLJ21955 fis, clone HEP05265. /FEA=mRNA /DB_XREF=gi:10438177 /UG=Hs.13255 KIAA0930 protein
218186 at	RAB25	NM_020387	1q21.2	CATX-8 protein (CATX-8)
211984 at		AI653730	14	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2
^211931 s at		BE867771	4	BX1
202040 s at	RBBP2	NM_005056	12p11	retinoblastoma-binding protein 2 (RBBP2)
214670 at		AA653300		Gessler Wilms tumor
213311 s at	KIAA1049	BF000251	16q24.3	KIAA1049 protein
^213015 at		BF448315	3	ESTs
213885 at	TRIM3	AA114843	11p15.5	tripartite motif-containing 3
201993 x at	HNRPDL	NM_005463	4q13-q21	heterogeneous nuclear ribonucleoprotein D-like (HNRPDL)
213729 at	FBNP3	Z78308	2q23.3	formin binding protein 3
213975 s at		AV711904	12	ESTs, Moderately similar to hypothetical protein FLJ23356
202305 s at	FEZ2	AI685892	2p21	fasciculation and elongation protein zeta 2 (zyglin II)
202309 at	MTHFD1	NM_005956	14q24	methylentetrahydrofolate dehydrogenase (NADP+ dependent)
202379 s at	NKTR	AI361805	3p23-p21	natural killer-tumor recognition sequence
218189 s at	SAS	NM_018946	9p24.1-p23	N-acetylneuraminic acid phosphate
201960 s at	KIAA0916	NM_015057	13q22	KIAA0916 protein
201959 s at	KIAA0916	AA488899	13q22	KIAA0916 protein

FIG. 1F

7/81

^212677 s at 214912 at 207618 s at ^208063 s at	RAB1A	AB0111154	2p14	Consensus includes gb:BG530481 /FEA=EST /DB_XREF=gi:13522018 /DB_XREF=est:602559046F1 /CLONE=IMAGE:4697218 /UG=Hs.79507 KIAA0582 protein FLJ12005 fis BCS1 (yeast homolog)-like (BCS1L)
	BCS1L	AK022067	2q33	
		NM_004328		
	CAPN9	NM_006615	1q42.11-q42.3	
222022 at				calpain 9 (nCL-4) (CAPN9)
		AW574818		Homo sapiens cDNA FLJ37066 fis, clone BRACE2015132, weakly similar to Drosophila melanogaster Oregon R cytoplasmic basic protein (deltex) mRNA
204205 at 208648 at	APOBEC3G	NM_021822	22q13.1-q13.2	gb:NM_021822.1 /DEF=Homo sapiens phorbol- like protein MDS019 (MDS019), mRNA. /FEA=mRNA /GEN=MDS019 /PROD=phorbol-like protein MDS019 /DB_XREF=gi:13399303 /UG=Hs.250619 phorbol-like protein MDS019 /FL=gb:AF182420.1 gb:NM_021822.1 valosin-containing protein
	VCP	W60953	9p13-p12	
220623 s at	TSGA10	NM_025244	2q11.2	gb:NM_025244.1 /DEF=Homo sapiens testis specific, 10 (TSGA10), mRNA. /FEA=mRNA /GEN=TSGA10 /PROD=testis specific, 10 /DB_XREF=gi:13376857 /UG=Hs.116116 testis specific, 10 /FL=gb:AF254756.1 gb:NM_025244.1
	DDX24	BG421209	14q32	
200702 s at 221419 s at		NM_013307		DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 24 non-functional folate binding protein

FIG. 1G

206104 at	ISL1	NM_002202	5q11.1	ISL1 transcription factor, LIMhomeodomain, (islet-1) (ISL1)
206438 x at	FLJ12975	NM_024809	12q24.31	hypothetical protein FLJ12975
91684 g at	FLJ20591	AI571298		exosome component Rrp41
206063 x at	PPIL2	BF001594	22q11.21	peptidylprolyl isomerase (cyclophilin)-like 2
204863 s at	IL6ST	BE856546	5q11	interleukin 6 signal transducer (gp130, oncostatin M receptor)
34063 at	RECQL5	AB006533	17q25.2-q25.3	RecQ protein-like 5
*200053 at	SPAG7	NM_004890	17p13.3	sperm associated antigen 7 (SPAG7)
44617 at	FLJ13491	AI431902	12q24.31	hypothetical protein FLJ13491
204026 s at	ZWINT	NM_007057	10q21-q22	ZW10 interactor (ZWINT)
201410 at		AI983043	2	ESTs
209579 s at	MBD4	AL556619	3q21-q22	methyl-CpG binding domain protein 4
209530 at	CACNB3	U07139	12q13	voltage-gated calcium channel beta subunit
201612 at	ALDH9A1	NM_000696	1q22-q23	aldehyde dehydrogenase 9 (gamma-aminobutyraldehyde dehydrogenase, E3 isozyme)
201596 x at	KRT18	NM_000224	12q13	keratin 18 (KRT18)
*218820 at	DKFZp761F2014	NM_020215	14q32.2	hypothetical protein DKFZp761F2014
219520 s at	KIAA1280	NM_018458	Xp22.32	uncharacterized bone marrow protein BM042
201137 s at	HLA-DPB1	NM_002121	6p21.3	major histocompatibility complex, class II, DP beta 1 (HLA-DPB1)
209056 s at	CDC5L	AW268817	6p21	CDC5 cell division cycle 5-like (S. pombe)
203707 at	ZNF263	NM_005741	16p13.11	zinc finger protein 263 (ZNF263)
219747 at	FLJ23191	NM_024574	4q27	hypothetical protein FLJ23191 (FLJ23191)
219458 s at	FLJ22609	NM_022072		hypothetical protein FLJ22609
59375 at	MYO15B	AI825877	17q25.3	myosin XVB, pseudogene

FIG. 1H

9/81

217022 s at	IGHM	S55735	14q32.33	immunoglobulin A1-A2 lambda hybrid GAU heavy chain VPS28 protein
211955 at	KPNB3	NM_002271	13q32.2	karyopherin (importin) beta 3 (KPNB3)
210042 s at	CTSZ	AF073890	20q13	cathepsin X precursor
^215295 at	DTNB	Y15718	2p23-p22	dystrobrevin B DTN-B2
216612 x at		AK021988		FLJ11926
213884 s at	TRIM3	AA114843	11p15.5	tripartite motif-containing 3
212651 at	RHOBTB1	AB018283	10q22.1	Consensus includes gb:AB018283.2 /DEF=Homo sapiens mRNA for KIAA0740 protein, partial cds. /FEA=mRNA /GEN=KIAA0740 /PROD=KIAA0740 protein /DB_XREF=gi:6705974 /UG=Hs.15099 KIAA0740 gene product
219816 s at	FLJ10482	NM_018107	14q11.1	gb:NM_018107.1 /DEF=Homo sapiens hypothetical protein FLJ10482 (FLJ10482), mRNA. /FEA=mRNA /GEN=FLJ10482 /PROD=hypothetical protein FLJ10482 /DB_XREF=gi:8922449 /UG=Hs.4997 hypothetical protein FLJ10482 /FL=gb:NM_018107.1
^208993 s at	PP1G	NM_004792	2q31.1	Consensus includes gb:AW340788 /FEA=EST /DB_XREF=gi:6837414 /DB_XREF=est:ha12h05.x1 /CLONE=IMAGE:2873529 /UG=Hs.77965 peptidyl-prolyl isomerase G (cyclophilin G) /FL=gb:U40763.1 gb:NM_004792.1
209119 x at	NR2F2	AV703465	15q26	nuclear receptor subfamily 2, group F, member 2

FIG. 1I

208994_s_at	PPIG	NM_004792	2q31.1	Consensus includes gb:AI638762 /FEA=EST /DB_XREF=gi:4690996 /DB_XREF=est:tt32a04.x1 /CLONE=IMAGE:2242446 /UG=Hs.77965 peptidyl-prolyl isomerase G (cyclophilin G) /FL=gb:U40763.1 gb:NM_004792.1
203904_x_at	KAI1	NM_002231	11p11.2	kangai 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2 leukocyte antigen, antigen detected by monoclonal and antibody IA4)) (KAI1)
220299_at	SPATA6	NM_019073	1p32.3	hypothetical protein (FLJ10007)
220038_at	SGKL	NM_013257	8q12.3-8q13.1	serumglucocorticoid regulated kinase-like (SGKL)
201088_at	KPNA2	NM_002266	17q23.1-q23.3	karyopherin alpha 2 (RAG cohort 1, importin alpha 1) (KPNA2)
208766_s_at	HNRP	BC001449	1p36.11	heterogeneous nuclear ribonucleoprotein R
219006_at	HSPC125	NM_014165	6q16.3	HSPC125 protein (HSPC125)
201419_at	BAP1	NM_004656	3p21.31-p21.2	BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase) (BAP1)
210088_x_at	MYL4	M36172	17q21-qter	embryonic myosin alkali light chain (MLC1)
209827_s_at	IL16	NM_004513	15q26.3	interleukin 16 (lymphocyte chemoattractant factor) (IL16)
219154_at		NM_024714		hypothetical protein FLJ12567
203663_s_at	COX5A	NM_004255	15q25	cytochrome c oxidase subunit Va (COX5A)
219455_at	FLJ21062	NM_024788	7p11.2	hypothetical protein FLJ21062
203632_s_at	GPRC5B	NM_016235	16p12	G protein-coupled receptor, family C, group 5, member B (GPRC5B)
219222_at	RBSK	NM_022128	2p23.3	ribokinase (RBSK)
^209205_s_at	LMO4	BC003600	1p22.3	transcription factor

FIG. 1J

11/81

204655_at	CCL5	NM_002985	17q11.2-q12	small inducible cytokine A5 (RANTES) (SCYA5)
36829_at	PER1	AF022991	17p13.1-17p12	period homolog 1 (Drosophila)
204373_s at	CAP350	NM_014810	1p36.13-q41	KIAA0480 gene product
222311_s at		AA648521	21	ESTs, Highly similar to SRA4_HUMAN CTD-binding SR-like protein RA4
*37012_at	CAPZB	U03271	1p36.1	capping protein (actin filament) muscle Z-line, beta
1405_i at	CCL5	M21121	17q11.2-q12	chemokine (C-C motif) ligand 5
^205831_at	CD2	NM_001767	1p13	CD2 antigen (p50), sheep red blood cell receptor (CD2)
205594_at	KIAA0924	NM_014897	17q22	KIAA0924 protein
206483_at	TSLRP	NM_012472	8q24.22	testis specific leucine rich repeat protein (TSLRP)
204144_s at	PIGQ	NM_004204	16p13.3	phosphatidylinositol glycan, class Q (PIGQ)
200914_x at	KTN1	BF589024	14q22.1	kinectin 1 (kinesin receptor)
220661_s at	FLJ20531	NM_017865	1q44	hypothetical protein FLJ20531
204033_at	TRIP13	NM_004237	5p15.2	thyroid hormone receptor interactor 13 (TRIP13)
208250_s at	DMBT1	NM_004406	10q25.3-q26.1	deleted in malignant brain tumors 1 (DMBT1), transcript variant 1
^200842_s at	EPRS	AI475965	1q41-q42	glutamyl-prolyl-tRNA synthetase
208093_s at	NUDEL	NM_030808	17p12	NUDEL; endooligopeptidase
221482_s at	ARPP-19	BC003418	15q15.3	cyclic AMP phosphoprotein
200887_s at	STAT1	NM_007315	2q32.2	signal transducer and activator of transcription 1, 91kD (STAT1)
202102_s at	BRD4	BF718610	19p13.1	bromodomain containing 4
212604_at	MRPS31	NM_005830	13q13.3	
214718_at	ODAG	AK026142	7q21-q22	hypothetical protein FLJ22489
213010_at	PRKCDBP	AI088622	11p15.4	protein kinase C, delta binding protein

FIG. 1K

12/81

215806 x at	TRGC2; TCRGC2; TRGC2(2X); TRGC2(3X)	M13231	7p15	T-cell receptor aberrantly rearranged gamma-chain
201687 s at	API5	NM_006595	11p12-q12	API5-like
^212226 s at	RPS20	AL576654	8q12	ribosomal protein S20
212208 at	KIAA1025	AK023837	12q24.21	thyroid hormone receptor-associated protein complex component TRAP240 mRNA.
212136 at	RPL35A	AW517686	3q29-qter	/FEA=mRNA /DB_XREF=gi:10435894 /UG=Hs.4084
214113 s at	RBM8A	AI738479	1q12	KIAA1025 protein
213465 s at	PPP1R7	BF718769	2q37.3	ribosomal protein L35a
^202018 s at	LTF	NM_002343	3q21-q23	RNA binding motif protein 8A
214055 x at	KIAA1096	AW238632	1q23.3	protein phosphatase 1, regulatory subunit 7
213392 at	MGC35048	AW070229	16p13.13	lactotransferrin (LTF)
213172 at	KIAA0227	D86980	14q24.1	KIAA1096 protein
214329 x at	TNFSF10	AV474434	3q26	hypothetical protein MGC35048
214337 at	COPA	AI621079	1q23-q25	KIAA0227 protein
201983 s at	EGFR	AW157070	7p12	tumor necrosis factor (ligand) superfamily, member 10
218150 at	ARL5	NM_012097	2q23.3	coatamer protein complex, subunit alpha
202688 at	TNFSF10	NM_003810	3q26	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)
202844 s at	RALBP1	AW025261	18p11.3	ADP-ribosylation factor-like 5 (ARL5)
211948 x at	KIAA1096	AL096857	1q23.3	tumor necrosis factor (ligand) superfamily, member 10
				ralA binding protein 1
				KIAA1096 protein

FIG. 1L



202585 s at	ODF2	NM_002504	9q34.11	nuclear transcription factor, X-box binding 1 (NFX1)
^210915 x at	IGHM	M15564	14q32.33	T-cell receptor rearranged beta-chain V-region (V-D-J)
210169 at	KIAA0420	AB007880	16p13.13	KIAA0420 mRNA
AFFX- HUMISGF3A/ M97935 3 at	STAT1	M97935	2q32.2	signal transducer and activator of transcription 1, 91kDa
217861 s at	PREB	NM_013388	2p23.3	prolactin regulatory element binding (PREB)
203318 s at	ZNF148	NM_021964	3q21	zinc finger protein 148 (pHZ-52) (ZNF148)
218286 s at	RNF7	NM_014245	3q22-q24	ring finger protein 7 (RNF7)
212052 s at	KIAA0676	AB014576	5q35.3	KIAA0676 protein
201671 x at	USP14	BC003556	18p11.32	ubiquitin specific protease 14 (tRNA-guanine transglycosylase)
201683 x at	KIAA0737	BE783632	14q11.1	KIAA0737 gene product
216469 at	PRED65	AL163202	21q11.1	putative gene with similarity to zinc finger proteins; Homo sapiens chromosome 21 segment HS21C002.
^216920 s at	TRG@	M27331	7p15-p14	T cell receptor gamma-chain
216976 s at	RYK	X96588	3q22	H-RYK receptor tyrosine kinase
216778 s at	CYLC1	Z22780	Xq21.1	cylicin, basic protein of sperm head cytoskeleton 1
215679 at		AL109706	15q24-q25	Consensus includes gb:AL109706.1 /DEF=Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 362430. /FEA=mRNA /DB_XREF=gi:5689823 /UG=Hs.32769 Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 362430

FIG. 1M

14/81

201877 s at	PPP2R5C	NM_002719	3p21	protein phosphatase 2, regulatory subunit B (B56), gamma isoform (PPP2R5C)
208448 x at	IFNA16	NM_002173	9p22	interferon, alpha 16 (IFNA16)
220724 at	FLJ21511	NM_025087	4p12	hypothetical protein FLJ21511 (FLJ21511)
205315 s at	SNTB2	NM_006750	16q22-q23	syntrophin, beta 2 (dystrophin-associated protein A1, 59kD, basic component 2) (SNTB2)
34260 at	KIAA0683	AB014583	16p13.12	KIAA0683 gene product

FIG. 1N

15/81

53003_analysis	Systematic	Common	Genbank	Map	Description
213217_at	ADCY2		AU149572	5p15.3	adenylate cyclase 2 (brain)
^213015_at			BF448315	3	ESTs
213703_at			W95043		FLJ33034 fis
#213441_x_at	PDEF		A1745526	6p21.3	prostate epithelium-specific Ets transcription factor
^212977_at	RDC1		A1817041	2q37.3	G protein-coupled receptor
	PTDSR; PSR; PTDSR1;				
212722_s_at	KIAA0585		AK021780	17q25	AK021780.1
^212677_s_at	RAB1A		AB011154	2p14	KIAA0582 protein
212933_x_at	RPL13		AA961748	16q24.3	ribosomal protein L13
212763_at	KIAA1078		AW593213	1q32.3	KIAA1078 protein
214858_at			AF070536		
^214685_at			AB000464	4p16.3	AB000464.1
^215295_at	DTNB		Y15718	2p23-p22	dystrobrevin B DTN-B2
215034_s_at	TM4SF1		A1189753	3q21-q25	transmembrane 4 superfamily member 1
214665_s_at			AK000095		calcium binding protein P22
^213851_at			BG031677	3	ESTs, Weakly similar to S65173 probable membrane protein YPL162c - yeast
213796_at	SPRR1A		A1923984	1q21-q22	small proline-rich protein 1A
#214404_x_at	PDEF		A1307915	6p21.3	prostate epithelium-specific Ets transcription factor
214164_x_at	FLJ20151		BF752277	15q21.3	hypothetical protein FLJ20151
210033_s_at	SPAG6		AF079363	10p12-p11.2	sperm flagellar protein Repro-SA-1 sperm associated antigen
#209921_at	SLC7A11		AB040875	4q28-q32	hxCIT mRNA for cystine/glutamate exchanger solute carrier family 7, (cationic amino acid transporter, y+ system) member 11
210445_at	FABP6		U19869	5q23-q35	fatty acid binding protein 6 (FABP6)
210037_s_at	NOS2A		L24553	17q11.2-q12	inducible nitric oxide synthase
209699_x_at	AKR1C1		U05598	10p15-p14	dihydrodiol dehydrogenase
209369_at	ANXA3		M63310	4q13-q22	1,2-cyclic-inositol-phosphate phosphodiesterase (ANX3)
#209249_s_at	GHITM		AF131820	10q23.31	
209644_x_at	CDKN2A		U38945	9p21	CDKN2A cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)

FIG. 2A

16/81

#209605_at	TST	D87292	22q13.1	rhodanese,
212399_s_at	GSA7	D50911	3p25.2	KIAA0121 protein
212354_at	KIAA1077	AW043713	8q13.1	sulfatase FP
212665_at	DKFZP434J214	AL556438	3q26.1	DKFZP434J214 protein
212493_s_at	HYPB	AI761110	3q22.3	Huntingtin interacting protein B
*212226_s_at	RPS20	AL576654	8q12	ribosomal protein S20
210854_x_at	SLC6A8	U17986	Xq28	GABAnoradrenaline transporter solute carrier family 6
3210524_x_at	RNAHP; RHELP; SF3b125	AF078844	17q22	hqp0376 RNA helicase-related protein
*211931_s_at		BE867771	4	Homo sapiens BX1 mRNA, partial cds
*210915_x_at	IGHM	M15564	14q32.33	T-cell receptor rearranged beta-chain V-region (V-D-J)
221726_at		BE250348		
221577_x_at	PLAB	AF003934	19p13.1-13.2	prostate differentiation factor
222016_s_at	ZNF323	AW086021	6	zinc finger protein 323
221775_x_at	RPL22	BG152979	1p36.3-p36.2	ribosomal protein L22
221253_s_at	MGC3178	NM_030810	6p25.2	hypothetical protein MGC3178
220192_x_at	PDEF	NM_012391	6p21.3	prostate epithelium-specific Ets transcription factor (PDEF)
219976_at	HOOK1	NM_015888	1p32.1	hook1 protein (HOOK1)
221059_s_at	CHST6	NM_021615	16q22	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6 (CHST6)
220389_at	FLJ23514	NM_021827	11q14.1	hypothetical protein FLJ23514 (FLJ23514)
44783_s_at	HEY1	R61374	8q21	hair/enhancer-of-split related with YRPW motif 1
38487_at	FLJ12442	D87433	3p21.31	hypothetical protein FLJ12442
823_at	CX3CL1	U84487	16q13	chemokine (C-X3-C motif) ligand 1
50221_at	TFEB	AI524138	6p21	transcription factor EB
37590_g_at		AL109698	15q24-q25	EST, Weakly similar to 2109260A B cell growth factor
222231_s_at	PRO1855	AK025328	17q23.3	highly similar to AF119857 hypothetical protein PRO1855
222122_s_at	THO2	BG403671	Xq25-q26.3	Tho2
33760_at	PEX14	AB017546	1p36.22	peroxisomal biogenesis factor 14
222288_at		AI004009		ESTs, Weakly similar to hypothetical protein FLJ20489

FIG. 2B

17/81

218142_s_at	TRNT1	NM_016302	3	LOC51185	
217853_at	TEM6	NM_022748	7p15.1	hypothetical protein FLJ13732 similar to tensin (FLJ13732)	
218647_s_at	FLJ23476	NM_024640	1p35.2	hypothetical protein FLJ23476	
218259_at	KIAA1243	NM_014048	16p13.2	gKIAA1243 protein	
217289_s_at	G6PC; G6PT; GSD1a	AF097831	17q21	glucose-6-phosphate transporter (G6PT) gene, G6PT-Dt allele	
3215785_s_at	CYFIP2	AL161999	5q34	DKFZp61H087 p53 inducible protein	
215755_at		AK022006		AK022006.1	
^216920_s_at	TRG@	M27331	7p15-p14	T cell receptor gamma-chain mRNA, C2 region	
^216072_at		AL050065	21q	DKFZp566M043	
#219156_at	SYNJ2BP	NM_018373	14q22.1-q24.3	hypothetical protein FLJ11271	
219118_at	FKBP11	NM_016594	12p11.1	FK506 binding protein precursor (LOC51303)	
219956_at	GALNT6	NM_007210	12q13	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) PTD013	
219598_s_at		NM_016104			
219117_s_at	FKBP11	NM_016594	12p11.1	FK506 binding protein precursor (LOC51303)	
218839_at	HEY1	NM_012258	8q21	hairyhancer-of-split related with YRPW motif 1 (HEY1)	
#218766_s_at	WARS2	NM_015836	1p13.3-p13.1	tryptophanyl tRNA synthetase 2 (mitochondrial) (WARS2), nuclear gene encoding mitochondrial protein	
218895_at	FLJ12455	NM_022078	1p35.3-p35.1	hypothetical protein FLJ12455	
218859_s_at	C20orf6	NM_016649	20p12.1	HDCMC28P protein	
203741_s_at	ADCY7	NM_001114	16q12-q13	adenylate cyclase 7 (ADCY7)	
203163_at	KATNB1	NM_005886	16q13	katanin p80 (WD40-containing) subunit B 1	
203767_s_at	STS	AU138166	Xp22.32	steroid sulfatase (microsomal), arylsulfatase C, isozyme S	
203755_at	BUB1B	NM_001211	15q15	budding uninhibited by benzimidazoles	
^203136_at	RABAC1	NM_006423	19q13.31	Rab acceptor 1 (prenylated) (RABAC1)	
#202747_s_at	ITM2A	NM_004867	Xq13.3-Xq21.2	integral membrane protein 2A (ITM2A)	
202672_s_at	ATF3	NM_001674	1q32.2	activating transcription factor 3 (ATF3)	
203026_at	KIAA0354	NM_014872	9p11.1	KIAA0354 gene product (KIAA0354)	
202766_s_at	FBN1	NM_000138	15q21.1	fibrillin 1 (Marfan syndrome) (FBN1)	
204612_at	PKIA	NM_006823	8q21.11	protein kinase (cAMP-dependent, catalytic) inhibitor alpha (PKIA)	

FIG. 2C

18/81

204604_at	PFTK1	NM_012395	7q21-q22	PFTK1
#204755_x_at	HLF	M95585	17q22	hepatic leukemia factor (HLF)
204754_at	HLF	A1810712	17q22	hepatic leukemia factor
#204351_at	S100P	NM_005980	4p16	S100 calcium-binding protein P (S100P)
203963_at	CA12	NM_001218	15q22	carbonic anhydrase XII (CA12)
203801_at	MRPS14	NM_022100	1q23-q25	hypothetical protein LOC63931
204179_at	MB	NM_005368	22q13.1	myoglobin (MB)
204160_s_at	ENPP4	AW194947	6p12.3	ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative function)
200908_s_at	RPLP2	BC005354	11p15.5-p15.4	ribosomal protein..
200872_at	S100A10	NM_002966	1q21	S100 calcium-binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11)) (S100A10)
201134_x_at	COX7C	NM_001867	5q14	cytochrome c oxidase subunit VIc (COX7C)
200986_at	SERPING1	NM_000062	11q12-q13.1	serine (or cysteine) proteinase inhibitor, clade G (C1 inhibitor)
#200842_s_at	EPRS	A1475965	1q41-q42	glutamyl-prolyl-IRNA synthetase
200762_at	DPYSL2	NM_001386	8p22-p21	dihydropyrimidinase-like 2 (DPYSL2)
200675_at	CD81	NM_004356	11p15.5	CD81 antigen (target of antiproliferative antibody 1)
200810_s_at	CIRBP	NM_001280	19p13.3	cold inducible RNA-binding protein (CIRBP)
200779_at	ATF4	NM_001675	22q13.1	activating transcription factor 4
202018_s_at	LTF	NM_002343	3q21-q23	lactotransferrin (LTF)
201898_s_at	UBE2A	A126625	Xq24-q25	ubiquitin-conjugating enzyme E2A (RAD6 homolog)
202528_at	GALE	NM_000403	1p36-p35	galactose-4-epimerase, UDP- (GALE)
#202437_s_at	CYP1B1	NM_000104	2p21	cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1 (glaucoma 3, primary infantile) (CYP1B1)
#201884_at	CEACAM5	NM_004363	19q13.1-q13.2	carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5)
201631_s_at	IER3	NM_003897	6p21.3	immediate early response 3 (IER3) immediate early response 3
201498_at	USP7	A160440	16p13.3	ubiquitin specific protease 7 (herpes virus-associated)
201709_s_at	NIPSNAP1	NM_003634	22q12.2	NIPSNAP, C. elegans, homolog 1 (NIPSNAP1)

FIG. 2D

19/81

201666_at	TIMP1	NM_003254	Xp11.3-p11.23	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)
200018_at	RPS13	NM_001017	11p15	ribosomal protein S13 (RPS13)
^208063_s_at	CAPN9	NM_006615	1q42.11-q42.3	calpain 9 (nCL-4) (CAPN9)
208047_s_at	NAB1	NM_005966	2q32.3-q33	NGF1-A binding protein 1 (ERG1 binding protein 1) (NAB1)
^208514_at	KCNE1	NM_000219	21q22.12	potassium voltage-gated channel, Isk-related family, member 1 (KCNE1)
208166_at	MMP16	NM_022564	8q21	matrix metalloproteinase 16 (membrane-inserted) (MMP16), transcript variant 2
207246_at	ZFY	NM_003411	Yp11.3	zinc finger protein, Y-linked (ZFY)
#207244_x_at	CYP2A6	NM_000762	19q13.2	cytochrome P450, subfamily 1A (phenobarbital-inducible)
207651_at	H963	NM_013308	3q26.1	platelet activating receptor homolog (H963)
207604_s_at	SLC4A7	NM_003615	3p22	solute carrier family 4, sodium bicarbonate cotransporter, member 7 (SLC4A7)
206884_s_at	SCEL	NM_003843	13q22	scellin (SCEL)
209125_at	KRT6A	J00269	12q12-q13	cytoskeletal type II keratin
209101_at	CTGF	M92934	6q23.1	connective tissue growth factor
^209205_s_at	LMO4	BC003600	1p22.3	LIM domain
209204_at	LMO4	A1824831	1p22.3	LIM domain
^208993_s_at	PPIG	NM_004792	2q31.1	peptidyl-prolyl isomerase G (cyclophilin G)
208819_at	MEL	BC002977	19p13.1	mel transforming oncogene (derived from cell line NK14)- RAB8 homolog
209031_at	IGSF4	NM_014333	11q23.2	immunoglobulin superfamily, member 4
209014_at	MAGED1	AF217963	Xp11.23	melanoma antigen, family D
208579_x_at	H2BFS	NM_017445	21q22.3	H2B histone family, member S (H2BFS)
#205566_at	ABHD2	NM_007011	15q26.1	putative transmembrane protein (HS1-2)
#205513_at	TCN1	NM_001062	11q11-q12	transcobalamin 1 (vitamin B12 binding protein, R binder family) (TCN1)
205726_at	DIAPH2	NM_006729	Xq22	diaphanous (Drosophila, homolog) 2 (DIAPH2), transcript variant 156
205680_at	MMP10	NM_002425	11q22.3	matrix metalloproteinase 10 (stromelysin 2) (MMP10)
#205328_at	CLDN10	NM_006984	13q31-q34	claudin 10 (CLDN10)

FIG. 2E

20/81

205239_at	AREG	NM_001657	4q13-q21	amphiregulin (schwannoma-derived growth factor) (AREG)
#205206_at	KAL1	NM_000216	Xp22.32	Kallmann syndrome 1 sequence (KAL1)
205319_at	PSCA	NM_005672	8q24.2	prostate stem cell antigen (PSCA)
#205278_at	GAD1	NM_000817	2q31	glutamate decarboxylase 1 (brain, 67kD) (GAD1), transcript variant GAD67
#205790_at	SCAP1	NM_003726	17q21.32	src kinase-associated phosphoprotein of 55 kDa (SKAP55)
^205831_at	CD2	NM_001767	1p13	CD2 antigen (p50), sheep red blood cell receptor (CD2)
205979_at	SCGB2A1	NM_002407	11q13	mammaglobin 2 (MGB2)
206561_s_at	AKR1B10	NM_020299	7q31.31	aldo-keto reductase family 1
205014_at	HBP17	NM_005130	4p16-p15	heparin-binding growth factor binding protein (HBP17)
205741_s_at	DTNA	NM_001392	18q12	dystrobrevin, alpha (DTNA)
#205759_s_at	SULT2B1	NM_004605	19q13.3	sulfotransferase family, cytosolic, 2B, member 1 (SULT2B1)
#205221_at	HGD	NM_000187	3q21-q23	homogentisate 1,2-dioxygenase (homogentisate oxidase) (HGD)
211126_s_at	CSRP2	U46006	12q21.1	smooth muscle LIM protein (h-SmLIM)
206276_at	E48	NM_003695	8q24-qter	lymphocyte antigen 6 complex, locus D (E48)
#219661_at	RANBP17	NM_022897	5q34	RAN binding protein 17 (RANBP17)

FIG. 2F



21/81

53003_analysis	Systematic	Common	Genbank	Map	Description
216346_at			AC004832		
219941_at	FLJ10936		NM_018279	12q13.3	hypothetical protein FLJ10936
202888_s_at	ANPEP		NM_001150	15q25-q26	alanine (membrane) aminopeptidase
212230_at	RPS20		AL576654	8q12	ribosomal protein S20
211710_x_at	RPL4		BC005817	15q22	ribosomal protein L4
212201_at	KIAA0692		AB014592	12	KIAA0692 protein
212126_at			BG391282	12	ESTs, Weakly similar to neuronal thread protein
213043_s_at	TRAP100		AI023317	17q12	thyroid hormone receptor-associated protein (100 kDa)
#213441_x_at	PDEF		AI745526	6p21.3	prostate epithelium-specific Ets transcription factor
213455_at	LOC92689		W87466	4p15.1	hypothetical protein BC001096
212270_x_at	RPL17		BG168283	18q21	ribosomal protein L17
212501_at	CEBPB		AL564683	20q13.1	CCAAT/enhancer binding protein (C/EBP), beta
212896_at	PPAP2A		D29641	5q11	KIAA0052 protein
209240_at			AF070560		O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)
#209249_s_at	GHITM		AF131820	10q23.31	Unknown
#209605_at	TST		D87292	22q13.1	rhodanese
208581_x_at	MT1L		NM_005952	16q13	metallothionein 1X (MT1X), mRNA
208864_s_at	TXN		AF313911	9q31	thioredoxin
208919_s_at	FLJ13052		BC001709	1p36.33-p36.21	Similar to hypothetical protein
210305_at	PDE4DIP		AB042557	1q12	similar to rat myomegalin
#210524_x_at	RNAHP; RHELP;		AF078844	17q22	hqp0376 protein helicase-related protein
211456_x_at	SF3b125		AF333388		metallothionein 1H-like protein
209737_at	AIP1		AB014605	7q21	KIAA0705 atrophin-1 interacting protein 1
209897_s_at	SLIT2		AF055585	4p15.2	neurogenic extracellular slit protein
218456_at	EEG1		NM_023925	12p11	hypothetical protein FLJ22569
*218820_at	DKFZp761F2014		NM_020215	14q32.2	hypothetical protein DKFZp761F2014
219792_at	AGMAT		NM_024758	1p36.13	hypothetical protein FLJ23384
217975_at	LOC51186		NM_016303	Xq22.1	pp21 homolog (LOC51186)
218175_at	FLJ22471		NM_025140	12q24.31	hypothetical protein FLJ22471
221841_s_at	KLF4		BF514079	9q31	Kruppel-like factor 4 (gut)
*37012_at	CAPZB		U03271	1p36.1	capping protein (actin filament) muscle Z-line, beta

FIG. 3A

22/81

43511_s_at		AI201594	11	ESTs
*218321_x_at	LOC51657	NM_016086	7p12.3	map kinase phosphatase-like protein MK-STYX (LOC51657)
221016_s_at	TCF-3	NM_031283	2p11.2	HMG-box transcription factor TCF-3 (TCF-3)
221030_s_at	DKFZP564B1162	NM_031305	4q21.23-q21.3	hypothetical protein DKFZP564B1162
#214404_x_at	PDEF	AI307915	6p21.3	prostate epithelium-specific Ets transcription factor
214736_s_at	ADD1	BE98639	4p16.3	adducin 1 (alpha)
215016_x_at	BPAG1	BC004912	6p12-p11	KIAA0728 Unknown protein
#209921_at	SLC7A11	AB040875	4q28-q32	hxCt mRNA for cystine/glutamate exchanger/ solute carrier family 7, (cationic amino acid transporter, y+ system)
213488_at	DKFZP566B2420	AL050143	2q37.3	DKFZP566B2420 protein
216336_x_at		AL031602		gene for a novel protein similar to MT1E (metallothionein 1E (functional)), ESTs, STSS, GSSs and two putative CpG island
217110_s_at	MUC4	AJ242547	3q29	sv7-MUC4 apomucin
217445_s_at	GART	AF008655	21q22.11	GARS-AIRS-GART
214033_at	ABCC6	AI084637	16p13.1	ATP-binding cassette, sub-family C (CFTR/MRP), member 6
215171_s_at	TIMM17A; TIM17;			FLJ13001 mitochondrial inner membrane preprotein translocase
#215785_s_at	TIM17A	AK023063	1q32.1	
	CYFIP2	AL161999	5q34	DKFZP761H087 p53 inducible protein
#202437_s_at	CYP1B1	NM_000104	2p21	cytochrome P450, subfamily 1 (dioxin-inducible), polypeptide 1 (glaucoma 3, primary infantile)
#202747_s_at	ITM2A	NM_004867	Xq13.3-Xq21.2	Integral membrane protein 2A (ITM2A)
203233_at	IL4R	NM_000418	16p11.2-12.1	interleukin 4 receptor (IL4R)
201722_s_at	GALNT1	NM_020474	18q12.1	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylglucosaminyltransferase 1 (GalNAc-T1)
*201833_at	HDAC2	NM_001527	6q21	histone deacetylase 2 (HDAC2)
#201884_at	CEACAM5	NM_004363	19q13.1-q13.2	carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5)
204059_s_at	ME1	NM_002395	6q12	malic enzyme 1, NADP(+)-dependent, cytosolic (ME1)
204341_at	TRIM16	NM_006470	17p11.2	estrogen-responsive B box protein (EBBP)
#204351_at	S100P	NM_005980	4p16	S100 calcium-binding protein P (S100P)
203476_at	TPBG	NM_006670	6q14-q15	5T4 oncofetal trophoblast glycoprotein (5T4)

FIG. 3B

23/81

203907_s_at	KIAA0763	NM_014869	3p25.1	KIAA0763 gene product
200062_s_at	RPL30	L05095	8q22	ribosomal protein L30
200940_s_at	REER	AB036737	1p36.1-p36.2	REER arginine-glutamic acid dipeptide (RE)
200972_at	TPAN-3	BC000704	15q23	tetraspan 3
1598_g_at	GAS6	L13720	13q34	growth arrest-specific 6
200034_s_at	RPL6	NM_000970	12q24.1	ribosomal protein L6 (RPL6)
*200053_at	SPAG7	NM_004890	17p13.3	sperm associated antigen 7 (SPAG7)
201431_s_at	DPYSL3	NM_001387	5q32	dihydropyrimidinase-like 3 (DPYSL3)
201442_s_at	ATP6IP2	AF248966	Xq21	HT028 mRNA ATPase, H+ transporting, lysosomal (vacuolar proton pump) membrane sector associated protein M8-9
203925_at	GCLM	NM_002061	1p22.1	glutamate-cysteine ligase, modifier subunit (GCLM)
201118_at	PGD	NM_002631	1p36.3-p36.13	phosphogluconate dehydrogenase (PGD)
201140_s_at	RAB5C	NM_004583	17q21.2	RAB5C, member RAS oncogene family
205789_at	CD1E	NM_001766	1q22-q23	CD1D antigen, d polypeptide
#205790_at	SCAP1	NM_003726	17q21.32	src kinase-associated phosphoprotein
205883_at	ZNF145	NM_006006	11q23.1	sapient zinc finger protein 145 (Kruppel-like, expressed in promyelocytic leukemia)
#205513_at	TCN1	NM_001062	11q11-q12	transcobalamin I (vitamin B12 binding protein, R binder family) (TCN1)
#205566_at	ABHD2	NM_007011	15q26.1	putative transmembrane protein (HS1-2)
205725_at	SCGB1A1	NM_003357	11q12.3-q13.1	uteroglobin (UGB)
206633_at	CHRNA1	NM_000079	2q24-q32	cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)
#207244_x_at	CYP2A6	NM_000762	19q13.2	cytochrome P450, subfamily IIA (phenobarbital-inducible), polypeptide 6
207547_s_at	TU3A	NM_007177	3p21.1	TU3A protein (TU3A)downregulated in renal cell carcinoma
206307_s_at	FOXDI	NM_004472	5q12-q13	forkhead box D1 (FOXDI)
206461_x_at	MT1H	NM_005951	16q13	metallothionein 1H (MT1H)
#204755_x_at	HLF	M95585	17q22	hepatic leukemia factor (HLF)
#205206_at	KAL1	NM_000216	Xp22.32	Kallmann syndrome 1 sequence (KAL1)
#205278_at	GAD1	NM_000817	2q31	glutamate decarboxylase 1 (brain, 67kD) (GAD1), transcript variant GAD67
201266_at	TXNRD1	NM_003330	12q23-q24.1	thioredoxin reductase 1
204505_s_at	EPB49	NM_001978	8p21.1	erythrocyte membrane protein band 4.9 (dematin)

FIG. 3C

24/81

204735_at	PDE4A	NM_006202	19p13.2	phosphodiesterase 4A, cAMP-specific
#205328_at	CLDN10	NM_006984	13q31-q34	claudin 10 (CLDN10)
205442_at	KIAA0626	NM_021647	4q32.3	KIAA0626 gene product
#206561_s_at	AKR1B10	NM_020299	7q31.31	aldo-keto reductase family 1, member B10
205499_at	SRPUL	NM_014467	Xq21.33-q23	sushi-repeat protein (SRPUL)
#219661_at	RANBP17	NM_022897	5q34	RAN binding protein 17 (RANBP17)
215844_at	TRN2	AK022217	19p13.2	FLJ12155
#205221_at	HGD	NM_000187	3q21-q23	homogentisate 1,2-dioxygenase
#205759_s_at	SULT2B1	NM_004605	19q13.3	sulfotransferase family, cytosolic, 2B, member 1

FIG. 3D

Fold change NS-S	Expression in Smoker compared to Non smoker	Fold change NS-SC	Expression in Smoker with cancer compared to Non smoker	Fold change S-SC	Expression in Smoker with cancer compared to smoker	Common	Genbank	Map	Chromosome	Description
1.308370044	UP	0.025522773	Down	0.0195073	Down	SFTPC	NM_003018	8p21	8	surfactant, pulmonary- associated protein C (SFTPC)
5.338061466	UP	0.69016811	Down	0.1292919	Down	FLJ21511	NM_025087	4p12	4	hypothetical protein FLJ21511 (FLJ21511)
1.400873362	UP	0.408836024	Down	0.29184367	Down		AL109706	15q24-q25	15	
7.578069671	UP	2.595974945	UP	0.34256414	Down	SCGB1A1	NM_003357	11q12.3-q13.1	11	uteroglobin (UGB)
0.635105608	Down	0.229272568	Down	0.36099912	Down		AL050065	21q	21	DKFZp566M043
1.21307779	UP	0.4400818	Down	0.36278119	Down	CYLC1	Z22780	Xq21.1	0	cylicin, basic protein of sperm head cytoskeleton 1
1.607085346	UP	0.595397727	Down	0.37048295	Down	PRKCDBP	A1088622	11p15.4	11	protein kinase C, delta binding protein

FIG. 4A

2.950396825	UP	1.156448477	UP	0.39196371	Down	PRED65	AL163202	21q11.1	21	putative gene with similarity to zinc finger proteins; Homo sapiens chromosome 21 segment HS21C002.
0.067702371	Down	0.029074395	Down	0.42944426	Down	AKR1B10	NM_020299	7q31.31	7	aldo-keto reductase family 1, member B10
2.904881101	UP	1.279375287	UP	0.4404226	Down	PDE4DIP	AB042557	1q12	1	similar to rat myomegalin
1.053114101	NC	0.464119455	Down	0.44071146	Down	CAPN9	NM_006615	1q42.11-q42.3	1	calpain 9 (nCL-4) (CAPN9)
3.049583151	UP	1.383543464	UP	0.45368281	Down	ANPEP	NM_001150	15q25-q26	15	alanine aminopeptidase (membrane)
1.52959706	UP	0.696038224	Down	0.45504678	Down		NM_024714		0	aminopeptidase (aminopeptidase N, M, microsomal aminopeptidase, CD13, p150) (ANPEP)
1.641799911	UP	0.762232351	Down	0.46426629	Down	FLJ12975	NM_024809	12q24.31	12	hypothetical protein FLJ12567
										hypothetical protein FLJ12975

FIG. 4B

1.458801051	UP	0.691341256	Down	0.47391058	Down	RYK	X96588	3q22	3	H-RYK receptor tyrosine kinase
1.603937464	UP	0.761512027	Down	0.47477663	Down	RHOB1B1	AB018283	10q22.1	10	KIAA0740 protein
2.89272944	UP	1.446077458	UP	0.4989007	Down	GPM6B	AF016004	Xp22.2	0	glycoprotein M6B
2.450994842	UP	1.269667568	UP	0.51802131	Down	DKFZp761F2014	NM_020215	14q32.2	14	hypothetical protein
1.570749543	UP	0.829744085	Down	0.52824722	Down	FLJ23191	NM_024574	4q27	4	hypothetical protein FLJ23191
2.118388503	UP	1.121950712	UP	0.52962462	Down	AK1	BC001116	9q34.1	9	adenylate kinase 1
1.673678691	UP	0.902488906	Down	0.53922439	Down	SPAG5	NM_006461	17q11.1	17	mitotic spindle coiled-coil related protein (DEEPST)
1.464681281	UP	0.82241624	Down	0.56149843	Down	RAGE	NM_014226	14q32	14	renal tumor antigen (RAGE)
0.812982556	Down	0.459165545	Down	0.56479139	Down	MYL4	M36172	17q21-qter	17	embryonic myosin alkali light chain (MLC1)
2.187452949	UP	1.238356813	UP	0.56611815	Down	RAB5C	NM_004583	17q21.2	17	RAB5C, member RAS oncogene family
0.203612479	Down	0.115767635	Down	0.56856846	Down	GAD1	NM_000817	2q31	2	glutamate decarboxylase 1
1.648155655	UP	0.938307692	NC	0.56930769	Down	RECQL5	AB006533	17q25.2-q25.3	17	RecQ protein-like 5

FIG. 4C

1.793250027	UP	1.038833852	NC	0.57930229	Down	TRIP13	NM_004237	5p15.2	5	thyroid hormone receptor interactor 13 (TRIP13)
1.15084901	UP	0.668732537	Down	0.58117856	Down	KIAA0420	AB007880	16p13.13	16	KIAA0420
1.503499079	UP	0.873933988	Down	0.58126673	Down	ODF2	NM_002504	9q34.11	9	nuclear transcription factor, X-box binding 1 (NFX1)
1.735468441	UP	1.016546543	NC	0.58574764	Down	MYO15B	AI825877	17q25.3	17	myosin XVB, pseudogene
1.638599046	UP	0.977624065	NC	0.59662189	Down	PER1	AF022991	17p13.1-17p12	17	period homolog 1 (Drosophila)
1.335571688	UP	0.805289075	Down	0.60295459	Down	TSGA10	NM_025244	2q11.2	2	testis specific, 10 (TSGA10)
1.26790138	UP	0.768995946	Down	0.60651085	Down	TRIM3	AA114843	11p15.5	11	tripartite motif- containing 3

FIG. 4D



0.835206368	Down	0.508488037	Down	0.60881724	Down	TIMP1	NM_003254	Xp11.3-p11.23	0	gb:NM_003254.1 /DEF=Homo sapiens tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor) (TIMP1), mRNA. /FEA=mRNA /GEN=TIMP1 /PROD=tissue inhibitor of metalloproteinase 1precursor /DB_XREF=gi:450 7508 /UG=Hs.5831 tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor) /FL=gb:BC000866 .1 gb:M12670.1 hypothetical protein FLJ21062
1.435207987	UP	0.875458421	Down	0.60998714	Down	FLJ21062	NM_024788	7p11.2	7	

FIG. 4E

0.537660505	Down	0.328592534	Down	0.61112972	Down	S100A10	NM_002966	1q21	1	S100 calcium-binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))
0.763479987	Down	0.469423729	Down	0.61484746	Down	DTNB	Y15718	2p23-p22	2	dystrobrevin B (p11)
1.497797357	UP	0.945041061	NC	0.63095389	Down	KIAA0683	AB014583	16p13.12	16	KIAA0683 gene product
1.402252011	UP	0.903454629	Down	0.64428835	Down	PP5395	NM_021732	10q24.32	10	hypothetical protein PP5395
1.264073638	UP	0.823674266	Down	0.65160307	Down	TSLRP	NM_012472	8q24.22	8	testis specific leucine rich repeat protein (TSLRP)
1.290263441	UP	0.842451316	Down	0.6529297	Down	FLJ10482	NM_018107	14q11.1	14	hypothetical protein FLJ10482
0.549316737	Down	0.367365405	Down	0.66876791	Down	CDKN2A	U38945	9p21	9	hypothetical 18.1 kDa protein (CDKN2A)
1.563056845	UP	1.047749895	NC	0.67032104	Down	RBM5	U23946	3p21.3	3	putative tumor suppressor (LUCA15)
1.1957413	UP	0.806776027	Down	0.67470784	Down	MGC35048	AW070229	16p13.13	16	hypothetical protein MGC35048

FIG. 4F

0.478899244	Down	0.326444589	Down	0.6816561	Down	PDEF	AI745526	6p21.3	6	prostate epithelium-specific Ets transcription factor
1.564614564	UP	1.067628032	NC	0.68235849	Down	LOC51657	NM_016086	7p12.3	7	map kinase phosphatase-like protein
1.872693415	UP	1.282798608	UP	0.68500193	Down	DKFZP586B2420	AL050143	2q37.3	2	DKFZP586B2420 protein
1.243906539	UP	0.860465116	Down	0.69174419	Down	PIGQ	NM_004204	16p13.3	16	phosphatidylinositol glycan, class Q (PIGQ)
2.557419355	UP	1.79863877	UP	0.70330224	Down	SLIT2	AF055585	4p15.2	4	neurogenic extracellular silt protein Slit2
0.918246857	NC	0.649737877	Down	0.70758519	Down	RABAC1	NM_006423	19q13.31	19	Rab acceptor 1 (prenylated) (RABAC1)
1.186212994	UP	0.850175043	Down	0.71072213	Down	RBSK	NM_022128	2p23.3	2	ribokinase (RBSK)
0.719208985	Down	0.514168065	Down	0.71490773	Down	DKFZP434J214	AL556438	3q26.1	3	DKFZP434J214 protein
1.160320592	UP	0.84103496	Down	0.72482982	Down		AI653730	14	14	Moderately similar to apolipoprotein B mRNA editing enzyme
1.308204606	UP	0.960529741	NC	0.7342351	Down	HNRPD1	D89092	4q13-q21	4	hnRNP J/K/TBP

FIG. 4G

0.948027169	NC	0.703663142	Down	0.74223943	Down	NUDEL	NM_030808	17p12	17	LIS1-interacting protein NUDEL; endooligopeptidase A (NUDEL)
1.329344184	UP	0.988549392	NC	0.7436369	Down	GRHPR; GLXR	AK026752	9q12	9	FLJ23099 fis, highly similar to AF113251 putative 2-hydroxyacid dehydrogenase mRNA
2.76724259	UP	2.059951203	UP	0.74440572	Down	CYFIP2	AL161999	5q34	5	DKFZp761H087 hypothetical protein p53 inducible protein
1.41097858	UP	1.051352347	NC	0.74512283	Down	ALDH2	NM_000690	12q24.2	12	aldehyde dehydrogenase 2, mitochondrial (ALDH2)
1.680427825	UP	1.263930929	UP	0.7521483	Down	GAS6	L13720	13q34	13	growth arrest-specific 6
1.286041667	UP	0.974555979	NC	0.75779503	Down	GBA	K02920	1q21	1	lysosomal glucocerebrosidase
1.652098093	UP	1.276821564	UP	0.77284852	Down	TCF-3	NM_031283	2p11.2	2	HMG-box transcription factor TCF-3
1.174029595	UP	0.912353309	NC	0.7771127	Down	USP4	AF017306	3p21.3	3	deubiquitinating enzyme UnpES

FIG. 4H

1.502337049	UP	1.167508233	UP	0.77712803	Down	SPAG7	NM_004890	17p13.3	17	sperm associated antigen 7 (SPAG7)
0.281762755	Down	0.221632397	Down	0.78659224	Down	CLDN10	NM_006984	13q31-q34	13	claudin 10 (CLDN10)
1.337478586	UP	1.063035804	NC	0.79480585	Down	PPAP2A	D29841	5q11	5	KIAA0052 protein
1.223830141	UP	0.98152846	NC	0.80201364	Down	HNRPD	NM_005463	4q13-q21	4	heterogeneous nuclear ribonucleoprotein D-like (HNRPD)
0.587868023	Down	0.473428695	Down	0.8053316	Down	ABHD2	NM_007011	15q26.1	15	putative transmembrane protein (HS1-2)
2.387529768	UP	2.001653726	UP	0.83837854	Down	KAL1	NM_000216	Xp22.32	0	Kallmann syndrome 1 sequence (KAL1)
0.494282334	Down	0.416042041	Down	0.84170931	Down	FKBP11	NM_016594	12p11.1	12	FK506 binding protein precursor (LOC51303)
1.43437574	UP	1.218311394	UP	0.84936698	Down		AF070560		0	O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)

FIG. 4I

0.53106691	Down	0.473970046	Down	0.8924865	Down	S100P	NIM_005980	4p16	4	S100 calcium-binding protein P (S100P)
1.598358991	UP	2.042740959	UP	1.27802388	Up	TST	D87292	22q13.1	22	rhodanese
1.612652412	UP	2.063668342	UP	1.27967337	Up	CX3CL1	U84487	16q13	16	chemokine (C-X3-C motif) ligand 1
0.520509757	Down	0.67025641	Down	1.28769231	Up	STS	AU138166	Xp22.32	0	steroid sulfatase (microsomal), arylsulfatase C, isozyme S
0.616465089	Down	0.805978739	Down	1.30737751	Up	FLJ13052	BC001709	1p36.33-p36.21	1	Similar to hypothetical protein FLJ13052
1.350159972	UP	1.795744681	UP	1.33002364	Up	RANBP17	NM_022897	5q34	5	RAN binding protein 17 (RANBP17)
0.153996101	Down	0.213033708	Down	1.38337079	Up	HGD	NM_000187	3q21-q23	3	homogentisate 1,2-dioxygenase (homogentisate oxidase) (HGD)
0.765723627	Down	1.073898737	NC	1.40246258	Up	CNP	BC001362	17q21	17	2,3-cyclic nucleotide 3 phosphodiesterase
1.447756051	UP	2.040716382	UP	1.40957199	Up	HEY1	NM_012258	8q21	8	hair/hemorrhoid-related split related with YRPW motif 1 (HEY1)

FIG. 4J

0.800213416	Down	1.180536673	UP	1.47527728	Up	VCP	W60953	9p13-p12	9	valosin-containing protein
0.949851543	NC	1.403407566	UP	1.47750201	Up		BE867771	4	4	BX1
0.916822921	NC	1.356213018	UP	1.47925296	Up	CDC5L	AW268817	6p21	6	CDC5 cell division cycle 5-like (S. pombe)
0.302546221	Down	0.453387342	Down	1.49857215	Up	SRPUL	NM_014467	Xq21.33-q23	0	sushi-repeat protein (SRPUL)
0.667153449	Down	1.004765262	NC	1.50604822	Up	HDAC2	NM_001527	6q21	6	histone deacetylase 2 (HDAC2)
1.225399013	UP	1.854924681	UP	1.51373117	Up	PFTK1	NM_012395	7q21-q22	7	PFTAIRE protein kinase 1 (PFTK1)
0.758070115	Down	1.189920545	UP	1.56967083	Up	RBM8A	AI738479	1q12	1	RNA binding motif protein 8A
0.671029785	Down	1.076956056	NC	1.60493033	Up	GS3955	NM_021643	2p25.1	2	
1.888361045	UP	3.07175713	UP	1.62667893	Up	LMO4	AI824831	1p22.3	1	LIM domain only 4
0.72418136	Down	1.192271442	UP	1.64637135	Up	KIAA1025	AK023837	12q24.21	12	similar to thyroid hormone receptor-associated protein complex component TRAP240
0.645745463	Down	1.069172451	NC	1.6557181	Up	RAD23B	AL527365	9q31.2	9	RAD23 homolog B (S. cerevisiae)

FIG. 4K

0.830530114	Down	1.437502002	UP	1.73082466	Up	PPP2R5C	NM_002719	3p21	3	phosphatase 2, regulatory subunit B (B56), gamma isoform (PPP2R5C)
0.663339495	Down	1.150213675	UP	1.73397436	Up	USP14	BC003556	18p11.32	18	ubiquitin specific protease 14 (tRNA-guanine transglycosylase)
0.735579896	Down	1.303463119	UP	1.77202113	Up	HNRPR	BC001449	1p36.11	1	heterogeneous nuclear ribonucleoprotein R
0.956569882	NC	1.701002735	UP	1.77823154	Up	EPRS	AI475965	1q41-q42	1	glutamyl-prolyl-tRNA synthetase
1.22220047	UP	2.220391227	UP	1.81671606	Up	RDC1	AI817041	2q37.3	2	G protein-coupled receptor
1.083865124	NC	1.983126177	UP	1.82967985	Up	RPS20	AL576654	8q12	8	ribosomal protein S20
0.676604315	Down	1.260724947	UP	1.86331201	Up	CAPZB	U03271	1p36.1	1	capping protein (actin filament) muscle Z-line, beta
0.699143688	Down	1.313625866	UP	1.87890685	Up	RPL35A	AW517686	3q29-qter	3	ribosomal protein L35a
0.662222645	Down	1.249595687	UP	1.88697215	Up	ARPP-19	BC003418	15q15.3	15	cyclic AMP phosphoprotein, ESTs
1.032214539	NC	1.955271565	UP	1.8942492	Up		BF448315	3	3	

FIG. 4L



0.9	NC	1.721052632	UP	1.9122807	Up	TRG@	M27331	7p15-p14	7	T cell receptor gamma-chain mRNA, C2 region
0.818202468	Down	1.580594832	UP	1.93178937	Up	MRPS31	NM_005830	13q13.3	13	
1.542148413	UP	3.036109674	UP	1.96875323	Up	FLJ23514	NM_021827	11q14.1	11	hypothetical protein FLJ23514
0.951531876	NC	1.932646707	UP	2.03108982	Up	ODAG	AK026142	7q21-q22	7	hypothetical protein FLJ22489
1.550558429	UP	3.167032406	UP	2.04251084	Up	LMO4	BC003600	1p22.3	1	LIM domain only 4
0.733895678	Down	1.526395866	UP	2.07985401	Up	TNFSF10	NM_003810	3q26	3	tumor necrosis factor (ligand) superfamily, member 10
0.577103432	Down	1.205490278	UP	2.08886347	Up	ACTR3	Z78330	2q13	2	ARP3 actin- related protein 3 homolog (yeast)
1.155659506	UP	2.41906807	UP	2.09323599	Up	RAB1A	AB011154	2p14	2	KIAA0582 protein
0.718102437	Down	1.51512374	UP	2.10989918	Up		AK022087		0	FLJ12005
1.032189203	NC	2.225802544	UP	2.15639007	Up	ZNF148	NM_021964	3q21	3	zinc finger protein 148 (pH2- 52)
0.710500191	Down	1.568640674	UP	2.20779788	Up		NM_013307		0	non-functional folate binding protein

FIG. 4M

0.698848053	Down	1.557351443	UP		2.22845501	Up	KHDRBS1	BC000717	1p32	1	GAP-associated tyrosine phosphoprotein p62 (Sam68)
0.67310203	Down	1.541355881	UP		2.28992904	Up	MBD4	AL556619	3q21-q22	3	methyl-CpG binding domain protein 4
2.24342816	UP	5.213389121	UP		2.32384937	Up	HLF	M95585	17q22	17	hepatic leukemia factor (HLF)
0.399851024	Down	0.98460018	NC		2.41239892	Up	MUC13	NM_017648	3q13.3	3	hypothetical protein FLJ20063
1.276552003	UP	3.154399758	UP		2.47103115	Up	EGFR	AW157070	7p12	7	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)
0.744773427	Down	1.844415243	UP		2.47647832	Up	SNTB2	NM_006750	16q22-q23	16	syntrophin, beta 2 (dystrophin-associated protein A1, 59kD, basic component 2)
0.777437905	Down	1.990289742	UP		2.56006265	Up	CD2	NM_001767	1p13	1	CD2 antigen (p50), sheep red blood cell receptor (CD2)

FIG. 4N

0.78687398	Down	2.04998135	UP	2.60522193	Up	VDU1	AK022864	1p31.1	1	FLJ12802 fis, clone NT2RP2002124, weakly similar to UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 4
0.751134301	Down	1.981105627	UP	2.61085884	Up	SYNJ2BP	NM_018373	14q22.1-q24.3	14	hypothetical protein FLJ11271
1.502381406	UP	4.103729402	UP	2.73148309	Up	DPYSL2	NM_001386	8p22-p21	8	dihydropyrimidina se-like 2 (DPYSL2)
0.61677674	Down	1.726232295	UP	2.79879603	Up	STAT1	M97935	2q32.2	2	signal transducer and activator of transcription 1, 91kDa
0.60633309	Down	1.757263088	UP	2.89818105	Up	STAT1	NM_007315	2q32.2	2	signal transducer and activator of transcription 1, 91kD (STAT1)
0.900203666	NC	2.657594937	UP	2.95221519	Up	KIAA0737	BE783632	14q11.1	14	KIAA0737 gene product
0.64696499	Down	2.119459088	UP	3.27600275	Up	KIAA1096	AL098857	1q23.3	1	KIAA1096 protein
0.692217573	Down	2.297068806	UP	3.31842024	Up	IGHM	M15564	14q32.33	14	T-cell receptor rearranged beta-chain V-region (V-D-J)

FIG. 40

0.809918754	Down	2.717463322	UP	3.35522953	Up	PPIG	NM_004792	2q31.1	2	peptidyl-prolyl isomerase G (cyclophilin G)
0.593513336	Down	2.143352601	UP	3.61129645	Up	DMBT1	NM_004406	10q25.3-q26.1	10	deleted in malignant brain tumors 1 (DMBT1)
0.612994291	Down	2.384544469	UP	3.88999458	Up	KTN1	BF589024	14q22.1	14	kinesin 1 (kinesin receptor)
0.515418408	Down	2.030847171	UP	3.94019139	Up	TNFSF10	AW474434	3q26	3	tumor necrosis factor (ligand) superfamily, member 10
0.425680325	Down	14.65960503	UP	34.438061	Up	LTF	NM_002343	3q21-q23	3	lactotransferrin (LTF)

FIG. 4P

41/81

53003_analysis	Common	Genbank	Map	Description
Systematic	GBA	K02920	1q21	lysosomal glucocerebrosidase
209093_s_at	RAGE	NM_014226	14q32	renal tumor antigen (RAGE)
219118_at	FKBP11	NM_016594	12p11.1	FK506 binding protein precursor (LOC51303)
203145_at	SPAG5	NM_006461	17q11.1	mitotic spindle coiled-coil related protein (DEEPEST)
216308_x_at	GRHPR; GLXR	AK026752	9q12	highly similar to AF113251 Homo sapiens putative 2-hydroxyacid dehydrogenase
214843_s_at	VDU1	AK022864	1p31.1	weakly similar to UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 4
AFFX- HUMISGF3A/M9793 5_3_at	STAT1	M97935	2q32.2	signal transducer and activator of transcription 1, 91kDa
208093_s_at	NUDEL	NM_030808	17p12	LIS1-interacting protein NUDEL;
213392_at	MGC35048	AW070229	16p13.13	endooligopeptidase A (NUDEL)
201488_x_at	KHDRBS1	BC000717	1p32	hypothetical protein MGC35048 GAP-associated tyrosine phosphoprotein p62 (Sam68)
211800_s_at	USP4	AF017306	3p21.3	deubiquitinating enzyme UnpES (UNP) mRNA, complete cds. ubiquitin specific protease 4 (proto-oncogene)
214912_at		AK022067		cDNA FLJ12005 fis
205328_at	CLDN10	NM_006984	13q31-q34	claudin 10 (CLDN10)
205206_at	KAL1	NM_000216	Xp22.32	Kallmann syndrome 1 sequence (KAL1)
201425_at	ALDH2	NM_000690	12q24.2	aldehyde dehydrogenase 2, mitochondrial (ALDH2)
219154_at		NM_024714		hypothetical protein FLJ12567
202478_at	GS3955	NM_021643	2p25.1	GS3955 protein (GS3955)
213102_at	ACTR3	Z78330	2q13	ARP3 actin-related protein 3 homolog (yeast)
201222_s_at	RAD23B	AL527365	9q31.2	RAD23 homolog B (S. cerevisiae)
209579_s_at	MBD4	AL556619	3q21-q22	methyl-CpG binding domain protein 4
201983_s_at	EGFR	AW157070	7p12	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)

FIG. 5A

42/81

215679_at		AL109706	15q24-q25	AL109706.1	heterogeneous nuclear ribonucleoprotein D-like (HNRPDL)
201993_x_at	HNRPDL	NM_005463	4q13-q21		
220389_at	FLJ23514	NM_021827	11q14.1		hypothetical protein FLJ23514
203318_s_at	ZNF148	NM_021964	3q21		zinc finger protein 148 (pHZ-52) (ZNF148)
200914_x_at	KTN1	BF589024	14q22.1		kinectin 1 (kinesin receptor)
208919_s_at	FLJ13052	BC001709	1p36.33-p36.21		Similar to hypothetical protein FLJ13052
206561_s_at	AKR1B10	NM_020299	7q31.31		aldo-keto reductase family 1, member B10 (aldose reductase) (AKR1B10)
213441_x_at	PDEF	A1745526	6p21.3		prostate epithelium-specific Ets transcription factor
220623_s_at	TSGA10	NM_025244	2q11.2		testis specific, 10 (TSGA10)
221419_s_at		NM_013307			non-functional folate binding protein (HSAF000381)
201877_s_at	PPP2R5C	NM_002719	3p21		protein phosphatase 2, regulatory subunit B (B56), gamma isoform (PPP2R5C)
212208_at	KIAA1025	AK023837	12q24.21		FLJ13775 fis, clone PLACE4000369, moderately similar to Homo sapiens thyroid hormone receptor-associated protein complex component TRAP240
213015_at		BF448315	3		ESTs
203136_at	RABAC1	NM_006423	19q13.31		Rab acceptor 1 (prenylated) (RABAC1)
200872_at	S100A10	NM_002966	1q21		S100 calcium-binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11)) (S100A10)
209240_at		AF070560			O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)
211931_s_at		BE867771	4		BX1
204351_at	S100P	NM_005980	4p16		S100 calcium-binding protein P (S100P)
205499_at	SRPUL	NM_014467	Xq21.33-q23		sushi-repeat protein (SRPUL)
202587_s_at	AK1	BC001116	9q34.1		adenylate kinase 1
210088_x_at	MYL4	M36172	17q21-qter		embryonic myosin alkali light chain (MLC1)
208063_s_at	CAPN9	NM_006615	1q42.11-q42.3		calpain 9 (nCL-4) (CAPN9)

FIG. 5B

43/81

202688_at	TNFSF10	NM_003810	3q26	tumor necrosis factor (ligand) superfamily, member 10 (TNFSF10)
204604_at	PFTK1	NM_012395	7q21-q22	PFTAIRE protein kinase 1 (PFTK1)
209067_s_at	HNRPD1	D89092	4q13-q21	hnRNP JKTBP mRNA heterogeneous nuclear ribonucleoprotein D-like
209644_x_at	CDKN2A	U38945	9p21	hypothetical 18.1 kDa protein (CDKN2A) cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
1598_g_at	GAS6	L13720	13q34	growth arrest-specific 6
209204_at	LMO4	AI824831	1p22.3	LIM domain only 4
211984_at		AI653730	14	ESTs, Moderately similar to apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2
823_at	CX3CL1	U84487	16q13	chemokine (C-X3-C motif) ligand 1
216778_s_at	CYLC1	Z22780	Xq21.1	cylicin, basic protein of sperm head cytoskeleton 1
203767_s_at	STS	AU138166	Xp22.32	steroid sulfatase (microsomal), arylsulfatase C, isozyme S
208250_s_at	DMBT1	NM_004406	10q25.3-q26.1	deleted in malignant brain tumors 1 (DMBT1), transcript variant 1
211948_x_at	KIAA1096	AL096857	1q23.3	KIAA1096 protein
200762_at	DPYSL2	NM_001386	8p22-p21	dihydropyrimidinase-like 2 (DPYSL2)
34063_at	RECQL5	AB006533	17q25.2-q25.3	RecQ protein-like 5
219747_at	FLJ23191	NM_024574	4q27	hypothetical protein FLJ23191
219816_s_at	FLJ10482	NM_018107	14q11.1	hypothetical protein FLJ10482
201683_x_at	KIAA0737	BE783632	14q11.1	KIAA0737 gene product
200887_s_at	STAT1	NM_007315	2q32.2	signal transducer and activator of transcription 1, 91kD (STAT1)
210169_at	KIAA0420	AB007880	16p13.13	KIAA0420
212977_at	RDC1	AI817041	2q37.3	G protein-coupled receptor
212226_s_at	RPS20	AL576654	8q12	ribosomal protein S20
216072_at		AL050065	21q	DKFZp566M043
209056_s_at	CDC5L	AW268817	6p21	CDC5 cell division cycle 5-like (S. pombe)
219156_at	SYNJ2BP	NM_018373	14q22.1-q24.3	hypothetical protein FLJ11271
218687_s_at	MUC13	NM_017648	3q13.3	hypothetical protein FLJ20063
218631_at	PP5395	NM_021732	10q24.32	hypothetical protein PP5395

FIG. 5C

44/81

212677_s_at	RAB1A	AB011154	2p14	KIAA0582 protein
201666_at	TIMP1	NM_003254	Xp11.3-p11.23	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor) (TIMP1)
201394_s_at	RBM5	U23946	3p21.3	putative tumor suppressor (LUCA15)
212665_at	DKFZP434J214	AL556438	3q26.1	DKFZP434J214 protein
215295_at	DTNB	Y15718	2p23-p22	dystrobrevin B DTN-B2
205982_x_at	SFTPC	NM_003018	8p21	surfactant, pulmonary-associated protein C (SFTPC)
200842_s_at	EPRS	A1475965	1q41-q42	glutamyl-prolyl-tRNA synthetase
220724_at	FLJ21511	NM_025087	4p12	hypothetical protein FLJ21511 (FLJ21511)
202888_s_at	ANPEP	NM_001150	15q25-q26	alanyl (membrane) aminopeptidase
218839_at	HEY1	NM_012258	8q21	haiyenhancer-of-split related with YRPW motif 1 (HEY1)
214718_at	ODAG	AK026142	7q21-q22	hypothetical protein FLJ22489
219661_at	RANBP17	NM_022897	5q34	RAN binding protein 17 (RANBP17)
208912_s_at	CNP	BC001362	17q21	2,3-cyclic nucleotide 3 phosphodiesterase
208766_s_at	HNRPR	BC001449	1p36.11	heterogeneous nuclear ribonucleoprotein R
208648_at	VCP	W60953	9p13-p12	valosin-containing protein
216920_s_at	TRG@	M27331	7p15-p14	T cell receptor gamma-chain mRNA, C2 region
210305_at	PDE4DIP	AB042557	1q12	similar to rat myomegalin
212896_at	PPAP2A	D29641	5q11	KIAA0052 protein
213885_at	TRIM3	AA114843	11p15.5	tripartite motif-containing 3
216976_s_at	RYK	X96588	3q22	H-RYK receptor tyrosine kinase
205831_at	CD2	NM_001767	1p13	CD2 antigen (p50), sheep red blood cell receptor (CD2)
205221_at	HGD	NM_000187	3q21-q23	homogentisate 1,2-dioxygenase (homogentisate oxidase) (HGD)
219222_at	RBSK	NM_022128	2p23.3	ribokinase (RBSK)
206483_at	TSLRP	NM_012472	8q24.22	testis specific leucine rich repeat protein (TSLRP)
219455_at	FLJ21062	NM_024788	7p11.2	hypothetical protein FLJ21062 (FLJ21062)
206438_x_at	FLJ12975	NM_024809	12q24.31	hypothetical protein FLJ12975
214329_x_at	TNFSF10	AW474434	3q26	tumor necrosis factor (ligand) superfamily, member 10

FIG. 5D



45/81

216469_at	PRED65	AL163202	21q11.1	putative gene with similarity to zinc finger proteins
204144_s_at	PIGQ	NM_004204	16p13.3	phosphatidylinositol glycan, class Q (PIGQ)
209605_at	TST	D87292	22q13.1	rhodanese, thiosulfate sulfurtransferase
212604_at	MRPS31	NM_005830	13q13.3	mitochondrion
208993_s_at	PIIG	NM_004792	2q31.1	peptidyl-prolyl isomerase G (cyclophilin G)
201140_s_at	RAB5C	NM_004583	17q21.2	RAB5C, member RAS oncogene family (RAB5C)
221016_s_at	TCF-3	NM_031283	2p11.2	HMG-box transcription factor TCF-3
209169_at	GPM6B	AF016004	Xp22.2	glycoprotein M6B
215785_s_at	CYFIP2	AL161999	5q34	DKFZp761H087 p53 inducible protein
205566_at	ABHD2	NM_007011	15q26.1	putative transmembrane protein (HS1-2)
205725_at	SCGB1A1	NM_003357	11q12.3-q13.1	utero globin (UGB)
212136_at	RPL35A	AW517686	3q29-qter	ribosomal protein L35a
209897_s_at	SLIT2	AF055585	4p15.2	neurogenic extracellular slit protein
36829_at	PER1	AF022991	17p13.1-17p12	period homolog 1 (Drosophila)
204033_at	TRIP13	NM_004237	5p15.2	thyroid hormone receptor interactor 13 (TRIP13)
204755_x_at	HLF	M95585	17q22	hepatic leukemia factor (HLF)
202585_s_at	ODF2	NM_002504	9q34.11	nuclear transcription factor, X-box binding 1 (NFX1)
210915_x_at	IGHM	M15564	14q32.33	T-cell receptor rearranged beta-chain V-region (V-D-J)
201833_at	HDAC2	NM_001527	6q21	histone deacetylase 2 (HDAC2)
59375_at	MYO15B	A1825877	17q25.3	myosin XVB, pseudogene
213488_at	DKFZP586B2420	AL050143	2q37.3	DKFZP586B2420 protein
212651_at	RHOB1	AB018283	10q22.1	KIAA0740 protein
201671_x_at	USP14	BC003556	18p11.32	ubiquitin specific protease 14 (tRNA-guanine transglycosylase)
214113_s_at	RBM8A	A1738479	11q12	RNA binding motif protein 8A
213010_at	PRKDCBP	A1088622	11p15.4	protein kinase C, delta binding protein
218321_x_at	LOC51657	NM_016086	7p12.3	MK-STYX (LOC51657)
202018_s_at	LTF	NM_002343	3q21-q23	lactotransferrin (LTF)
221482_s_at	ARPP-19	BC003418	15q15.3	cyclic AMP phosphoprotein
205278_at	GAD1	NM_000817	2q31	glutamate decarboxylase 1 (brain, 67kD) (GAD1)

FIG. 5B

46/81

34260_at	KIAA0683	AB014583	16p13.12	KIAA0683 gene product
218820_at	DKFZp761F2014	NM_020215	14q32.2	hypothetical protein DKFZp761F2014
200053_at	SPAG7	NM_004890	17p13.3	sperm associated antigen 7 (SPAG7)
205315_s_at	SNTB2	NM_006750	16q22-q23	syntrophin, beta 2 (dystrophin-associated protein)
37012_at	CAPZB	U03271	1p36.1	capping protein (actin filament) muscle Z-line, beta
209205_s_at	LMO4	BC003600	1p22.3	LIM domain only 4

FIG. 5F

47/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_P Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
209074_s_at		AL050264. 1	0.00000001	Down	0.25762095	TU3A protein matrix metalloproteinase 10 (stromelysin 2)	NULL	3p21.1	11170	Hs.8022	
205680_at	MMP10	NM_00242 5.1	0.00000004	Down	0.39706706		collagen catabolism // inferred from electronic annotation	11q22.3	4319	Hs.2258	185260
204755_x_at	HLF	M95585.1 0	0	Down	0.4701625	hepatic leukemia factor	regulation of transcription, DNA- dependent // inferred from electronic annotation	17q22	3131	Hs.250692	142385
210524_x_at		AF078844. 1	0.000000781	Down	0.52723124	RNA helicase-related protein	cytoplasm // non- traceable author statement	17q23	11325	Hs.8765	
203687_at	SCYD1	NM_00299 6.1	0.00000006	Down	0.53669865	small inducible cytokine subfamily D (Cys-X3-Cys), member 1 (fractalkine, neurotactin)	immune response // not recorded	16q13	6376	Hs.80420	601880
204379_s_at	FGFR3	NM_00014 2.2	0.00003694	Down	0.54726067	fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)	skeletal development // predicted/computed	4p16.3	2261	Hs.1420	134934
206461_x_at	MT1H	NM_00595 1.1	0.00002215	Down	0.55093068	metallothionein 1H	metal ion binding // inferred from electronic annotation	16q13	4496	Hs.2667	156354
208561_x_at	MT1X	NM_00595 2.1	0.00000031	Down	0.55320327	metallothionein 1X	heavy metal sensitivity/resistance // traceable author statement	16q13	4501	Hs.278462	156359

FIG. 6A

48/81

Gene list BU.AFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
823_at	SCYD1	U84487	0	Down	0.55735498	small inducible cytokine subfamily D (Cys-X3-Cys), member 1 (fractalkine, neurotactin)	immune response // not recorded	16q13	6376	Hs.80420	601880
204754_at	HLF	AI810712	0.00000017	Down	0.56297333	hepatic leukemia factor	regulation of transcription, DNA- dependent // inferred from electronic annotation	17q22	3131	Hs.250692	142385
213629_x_at		BF246115	0.00000001	Down	0.56721444	RNA helicase-related protein	cytoplasm // non- traceable author statement				
203914_x_at	HPGD	NM_00086 0.1	0.00003925	Down	0.56944096	hydroxyprostaglandin dehydrogenase 15- (NAD)	metabolism // Inferred from electronic annotation	4q34-q35	3248	Hs.77348	601688
205206_at	KAL1	NM_00021 6.1	0.00003575	Down	0.58642444	Kallmann syndrome 1 sequence	cell motility // traceable author statement	xp22.32	3730	Hs.89591	308700
209897_s_at	SLIT2	AF055585. 1	0.00001259	Down	0.58826137	slit homolog 2 (Drosophila)	mesoderm migration // inferred from mutant phenotype	4p15.2	9353	Hs.29802	603746
200878_at	EPAS1	AF052094. 1	0.00000045	Down	0.59071696		NULL	2p21-p16	2034	Hs.8136	603349
211538_s_at		U56725.1	0.00000016	Down	0.59469726	heat shock 70kD protein 2	spermatid development // traceable author statement			Hs.75452	

FIG. 6B

49/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS moker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
204745_x_at	MT1G	NM_00595 0.1	0.00000147	Down	0.60076773	metallothionein 1G	metal ion binding // inferred from electronic annotation	16q13	4485	Hs.334409	156353
202746_at	ITM2A	AL021786	0.00005283	Down	0.60146828		integral to membrane // inferred from electronic annotation	xq13.3- xq21.2	9452	Hs.17109	300222
215785_s_at		AL161999. 1	0.00000053	Down	0.60939296	cytoplasmic FMR1 interacting protein 2	NULL			Hs.258503	
217165_x_at	MT1F	M10943	0.00000031	Down	0.61734582		cytoplasm // non- traceable author statement	16q13	4494	Hs.203936	156352
204326_x_at	MT1L	NM_00245 0.1	0.00000002	Down	0.61891557	serine threonine kinase 39 (STE20/SPS1 homolog, yeast)	cytoplasm // non- traceable author statement	16q13	4500	Hs.94360	156358
209784_s_at	JAG2	AF029778. 1	0.00002692	Down	0.63262088	jagged 2	N receptor signaling pathway // non- traceable author statement	14q32	3714	Hs.166154	602570
220389_at		NM_02182 7.1	0.00000047	Down	0.66833583	hypothetical protein FLJ23514	NULL	11q14.1	60494	Hs.144913	
213601_at	SLIT1	AB011537. 2	0.00000089	Down	0.67196823	slit homolog 1 (Drosophila)	NULL	10q23.3- q24	6585	Hs.133466	603742
200953_s_at		NM_00175 9.1	0.00000909	Down	0.6749563	cyclin D2	cytokinesis // inferred from electronic annotation				
221748_s_at		AL046979	0.00006342	Down	0.68035154	Homo sapiens cDNA FLJ32766 fis, clone TEST12001862	NULL			Hs.9973	

FIG. 6C

50/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
201397_at	PHGDH	NM_00662 3.1	0.00004514	Down	0.68087588	phosphoglycerate dehydrogenase	serine biosynthesis // not recorded	1p12	26227	Hs.3343	601815
213169_at						Homo sapiens cDNA FLJ12815 fls, clone NT2RP2002546	NULL				
1598_g_at	GAS6	BG109855 L13720	0.00000668	Down	0.68210027	growth arrest-specific 6	NULL	13q34	2621	Hs.78501	600441
221016_s_at						HMG-box transcription factor TCF-3	regulation of transcription, DNA- dependent // non- traceable author statement	2p11.2	83439	Hs.102367	
212185_x_at	MT2A	NM_00595 3.1	0.00000019	Down	0.69786031		copper ion homeostasis // traceable author statement	16q13	4502	Hs.118786	156360
32137_at	JAG2	AF029778	0.0000334	Down	0.69985468	metallothionein 2A jagged 2	N receptor signaling pathway // non- traceable author statement	14q32	3714	Hs.166154	602570
213029_at						Homo sapiens mRNA; cDNA DKFZp564H1916 (from clone DKFZp564H1916)	NULL				
212126_at						Homo sapiens cDNA FLJ31079 fls, clone HSYRA2001595	NULL			Hs.350401	

FIG. 6D

51/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS moker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
207760_s_at	NCOR2	NM_00631 2.1	0.00003465	Down	0.74126481	nuclear receptor co- repressor 2	DNA binding activity // extended: unknown; myb_DNA-binding; 3.9e- 10	12q24	9612	Hs.287994	600848
200675_at	CD81	NM_00435 6.1	0.00000198	Down	0.74141757	CD81 antigen (target of antiproliferative antibody 1)	protein complex assembly // not recorded	11p15	975	Hs.54457	186845
212226_s_at		AL576654	0.00006297	Down	0.7547925	phosphatidic acid phosphatase type 2B	NULL				
211026_s_at	MGLL	BC006230. 1	0.00000672	Down	0.75734151	monoglyceride lipase	inflammatory response // traceable author statement	3q21.3	11343	Hs.6721	
203407_at	PPL	NM_00270 5.1	0.0000534	Down	0.79282943	periplakin	cytoskeleton // traceable author statement	16p13.3	5493	Hs.74304	602871
200642_at	SOD1	NM_00045 4.1	0.00002194	Up	1.23027558	superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult))	response to oxidative stress // not recorded	21q22.11	6647	Hs.75428	147450
208799_at	PSMB5	BC004146. 1	0.00003406	Up	1.23472792	proteasome (prosome, macropain) subunit, beta type, 5	NULL				
212181_s_at		AF191654. 2	0.00003238	Up	1.24764073		cyclic nucleotide metabolism // experimental evidence				

FIG. 6E

52/81

Gene list BU AFFX_ID	GENE_SYM	GB_ACC	Smoker_NS maker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category //	MAP	LOCUS LINK	UNIGENE	OMIM
208689_s_at	RPN2	BC003560. 1	0.00002416	Up	1.26443482	ribophorin II	protein modification // traceable author statement	20q12- q13.1	6185	Hs.75722	180490
200599_s_at	TRA1	NM_00329 9.1	0.0000016	Up	1.28607041	tumor rejection antigen (gp96) 1	ATP binding activity // extended:Unknown; HATPase_c; 5.4e-13	12q24.2- q24.3	7184	Hs.82689	191175
200863_s_at	RAB11A	A1215102	0.0000151	Up	1.28734777	RAB11A, member RAS oncogene family	intracellular protein transport // inferred from electronic annotation	15q21.3- q22.31	8766	Hs.75618	605570
209694_at	PTS	M97655.1 NM_00667	0.00003398	Up	1.29325473	6- pyruvoy/tetrahydropter- in synthase	central nervous system development // predicted/computed	11q22.3- q23.3	5805	Hs.366	261640
203476_at	TPBG	0.1	0.00002117	Up	1.30815003	trophoblast glycoprotein	cell motility // not recorded	6q14-q15	7162	Hs.82128	190920
201923_at	PRDX4	NM_00640 6.1	0.00001455	Up	1.31811994	peroxiredoxin 4	I-kappaB phosphorylation // traceable author statement	xp22.13	10549	Hs.83383	606506
203509_at	SORL1	NM_00310 5.2	0.00002403	Up	1.34981345	sortilin-related receptor, L(DLR class) A repeats- containing	intracellular protein transport // inferred from electronic annotation	11q23.2- q24.2	6653	Hs.278571	602005
208918_s_at		BC001709. 1	0.00001625	Up	1.35785089	NAD kinase	NULL			Hs.220324	
205566_at		NM_00701 1.1	0.00004071	Up	1.36323619	lung alpha/beta hydrolase 2	integral to membrane // non-traceable author statement	15q28.1	11057	Hs.99364	

FIG. 6F



53/81

Gene list BU_AFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
202804_at	ABCC1	AI539710	0.00005085	Up	1.36862413	ATP-binding cassette, sub-family C (CFTR/MRP), member 1	small molecule transport // traceable author statement	16p13.1	4363	Hs.89433	158343
200967_at	PPIB	NM_00094 2.1	0.00004739	Up	1.37577594	peptidylprolyl isomerase B (cyclophilin B)	protein folding // inferred from electronic annotation	15q21-q22	5479	Hs.699	123841
217955_at		NM_01536 7.1	0.00005905	Up	1.37660816	MIL1 protein	caspase activation // non-traceable author statement	22q11	23786	Hs.10267	
208919_s_at		BC001709. 1	0.00000238	Up	1.37784082	NAD kinase	NULL			Hs.220324	
202481_at		NM_00475 3.1	0.00000325	Up	1.38649421	short-chain dehydrogenase/reduc tase 1	metabolism // inferred from electronic annotation	1p36.1	9249	Hs.17144	
201272_at	AKR1B1	NM_00162 8.1	0.0000218	Up	1.4007013	aldo-keto reductase family 1, member B1 (aldose reductase)	carbohydrate metabolism // traceable author statement	7q35	231	Hs.75313	103880
207222_at	PLA2G10	NM_00356 1.1	0.00002852	Up	1.42588829	phospholipase A2, group X	lipid catabolism // inferred from electronic annotation	16p13.1- p12	8399	Hs.144442	603603
204500_s_at		NM_01523 9.1	0.00001725	Up	1.43452772	ATP/GTP binding protein 1	NULL	9q21.32	23287	Hs.21542	
209114_at		AF133425. 1	0.00000494	Up	1.4367707	tetraspan 1	cell proliferation // not recorded	1pter- p32.1	10103	Hs.38972	
206302_s_at	NUDT4	NM_01909 4.1	0.00000098	Up	1.43822734	nudix (nucleoside diphosphate linked moiety X)-type motif 4	cyclic nucleotide metabolism // experimental evidence	12q21	11163	Hs.92381	

FIG. 6G

54/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
213059_at		AF055009.1	0.00000024	Up	1.47009393		regulation of transcription, DNA-dependent // inferred from electronic annotation"				
206094_x_at	UGT1A6	NM_001072.1	0.00002939	Up	1.47638518	UDP glycosyltransferase 1 family, polypeptide A6	xenobiotic metabolism // inferred from direct assay	2q37	54578	Hs.284239	606431
200748_s_at	FTH1	NM_002032.1	0	Up	1.48230063	ferritin, heavy polypeptide 1	iron ion transport // inferred from electronic annotation	11q13	2495	Hs.62954	134770
213455_at		W87466	0.00000003	Up	1.4858762	Homo sapiens cDNA FLJ33151 fis, clone UTERU2000263	NULL			Hs.246885	
213624_at		AA873600	0.00005778	Up	1.49230544	acid sphingomyelinase-like phosphodiesterase	hydrolase activity // extended:Unknown; Metallophos; 1.3e-09				
205042_at		NM_005476.2	0.0000345	Up	1.49380148	UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase	N-acetylneuraminat metabolism // traceable author statement	9p11.2	10020	Hs.5920	603824
203397_s_at	GALNT3	BF063271	0.00001254	Up	1.49452656	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 3 (GalNAcT3)	carbohydrate metabolism // not recorded	2q24-q31	2591	Hs.278611	601756

FIG. 6H

55/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
222016_s_at		AW086021	0.00002754	Up	1.50397769	hypothetical protein FLJ23407	regulation of transcription, DNA- dependent // inferred from electronic annotation"	6p21.32	79764	Hs.104842	
203126_at	IMPA2	NM_01421 4.1	0.00002945	Up	1.50470629	inositol(myo)-1(or 4)- monophosphatase 2	signal transduction // traceable author statement	18p11.2	3613	Hs.5753	605922
214105_s_at		A1762113	0.00000069	Up	1.5059961	GDP-mannose 4,6- dehydratase	carbohydrate metabolism // traceable author statement			Hs.105435	
201263_at	TARS	NM_00319 1.1	0.00002058	Up	1.5344928	threonyl-tRNA synthetase	threonyl-tRNA aminoacylation // inferred from electronic annotation	5p13-can	6897	Hs.84131	187780
204532_x_at	UGT1A9	NM_02102 7.1	0.0000014	Up	1.54255992	UDP glycosyltransferase 1 family, polypeptide A9	metabolism // traceable author statement	2q37	54600	Hs.2056	606434
209448_at	HTATIP2	BC002439. 1	0.00000843	Up	1.54504237	HIV-1 Tat interactive protein 2, 30 kD	induction of apoptosis // experimental evidence	11p15.1	10553	Hs.90753	605628
204970_s_at	MAFG	NM_00235 9.1	0.00001062	Up	1.54913013	v-maf musculoaponeurotic fibrosarcoma oncogene homolog G (avian)	regulation of transcription, DNA- dependent // inferred from electronic annotation	17q25	4097	Hs.252229	602020

FIG. 6I

56/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
200872_at	S100A10	NM_00296 6.1	0.00001827	Up	1.59972573	S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))	signal transduction // not recorded	1q21	6281	Hs.119301	114085
208596_s_at	UGT1A3	NM_01909 3.1	0.00003102	Up	1.61820133	UDP glycosyltransferase 1 family, polypeptide A3	metabolism // traceable author statement	2q37	54659	Hs.326543	606428
208680_at	PRDX1	L19184.1	0.00000042	Up	1.62489119	peroxiredoxin 1	cell proliferation // traceable author statement	1p34.1	5052	Hs.180909	176763
215867_x_at		AL050025. 1	0.00000108	Up	1.62599641	adaptor-related protein complex 1, gamma 1 subunit	endocytosis // inferred from electronic annotation			Hs.5344	
218313_s_at	GALNT7	NM_01742 3.1	0.00000011	Up	1.63625819	UDP-N-acetyl-alpha- D- galactosamine:polype- ptide N- acetylglactosaminyl/ transferase 7 (GalNAc T7)	carbohydrate metabolism // traceable author statement	4q31.1	51809	Hs.246315	605005
209101_at	CTGF	M52934.1	0.00005544	Up	1.65588494	connective tissue growth factor	regulation of cell growth // inferred from electronic annotation	6q23.1	1490	Hs.75511	121009
214211_at		AA083483	0.00002263	Up	1.6618808	ferritin, heavy polypeptide 1	NULL			Hs.62954	

FIG. 6J

57/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS moker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
201487_at	CTSC	NM_00181 4.1	0.00006587	Up	1.66952894	cathepsin C	proteolysis and peptidolysis // not recorded	11q14.1- q14.3	1075	Hs.10029	602365
209173_at	AGR2	AF088867 1	0.00000667	Up	1.66783185	anterior gradient 2 homolog (Xenopus laevis)	oncogenesis // traceable author statement	7p21.3	10551	Hs.91011	606358
207126_x_at	UGT1A1	NM_00046 3.1	0.00000129	Up	1.69648422	UDP glycosyltransferase 1 family, polypeptide A1	transferase activity, transferring hexosyl groups // extended:Unknown; UDPGT; 0	2q37	54658	Hs.278898	191740
201431_s_at	DPYSL3	NM_00138 7.1	0.00000005	Up	1.72880024	dihydropyrimidinase- like 3	signal transduction // traceable author statement	5q32	1809	Hs.74566	601168
217755_at	HN1	NM_01618 5.1	0.00000002	Up	1.73204629	hematological and neurological expressed 1	NULL	17q25	51155	Hs.109706	
203963_at	CA12	NM_00121 8.2	0.00000054	Up	1.75150463	carbonic anhydrase XII	one-carbon compound metabolism // inferred from electronic annotation	15q22	771	Hs.5338	603263
202923_s_at	GCLC	NM_00149 8.1	0.00000017	Up	1.77328052	glutamate-cysteine ligase, catalytic subunit	cysteine metabolism // traceable author statement	6p12	2729	Hs.151393	230450
208700_s_at	TKT	L12711.1	0.00000126	Up	1.80682503	transketolase (Wernicke-Korsakoff syndrome)	transferase activity // inferred from electronic annotation	3p14.3	7086	Hs.89843	277730
204875_s_at	GMDS	NM_00150 0.1	0.00000008	Up	1.83056948	GDP-mannose 4,6- dehydratase	carbohydrate metabolism // traceable author statement	6p25	2762	Hs.105435	602884

FIG. 6K

58/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
219928_s_at		NM_01218 9.1	0.00001001	Up	1.83997341	calcium-binding tyrosine-(Y)- phosphorylation regulated (fibrous sheathin 2)	cAMP-dependent protein kinase, regulator activity // extended: inferred from electronic annotation; Rlla; 4.5e-15	18p11.1	26256	Hs.314452	
203824_at	TM4SF3	NM_00461 6.1	0.000000093	Up	1.8579966	transmembrane 4 superfamily member 3	pathogenesis // traceable author statement	12q14.1- q21.1	7103	Hs.84072	600769
201266_at	TXNRD1	NM_00333 0.1	0.000000003	Up	1.8650579	thioredoxin reductase 1	electron transport // inferred from electronic annotation	12q23- q24.1	7296	Hs.13046	601112
201118_at	PGD	NM_00263 1.1	0.000000023	Up	1.86620672	phosphogluconate dehydrogenase	pentose-phosphate shunt, oxidative branch // traceable author statement"	1p36.3- p36.13	5226	Hs.75888	172200
209369_at	ANXA3	M63310.1	0.000000025	Up	1.8728617	annexin A3	cytoplasm // not recorded	4q13-q22	306	Hs.1378	106490
203925_at	GCLM	NM_00206 1.1	0.00000154	Up	1.87522013	glutamate-cysteine ligase, modifier subunit	cysteine metabolism // traceable author statement	1p22.1	2730	Hs.89709	601176
214858_at		AF070536. 1	0.00001621	Up	1.9180029		NULL				
211657_at		M18728.1	0.000000002	Up	1.92577464		signal transduction // traceable author statement				
210065_s_at	UPK1B	AB002155. 1	0.000003154	Up	1.95451447	uroplakin 1B	physiological processes // non-traceable author statement	3q13.3- q21	7348	Hs.271580	602380

FIG. 6L

59/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS moker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category //	MAP	LOCUS LINK	UNIGENE	OMIM
208864_s_at	TXN	AF313911. 1	0	Up	1.96132242	thioredoxin	signal transduction //	9q31	7295	Hs.76136	187700
201463_s_at	TALDO1	NM_00675 5.1	0	Up	1.97483909	transaldolase 1	carbohydrate metabolism // traceable author statement	11p15.5- p15.4	6888	Hs.77290	602063
203757_s_at	CEACAM6	BC005008. 1	0.00000165	Up	1.98733622	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	signal transduction //	19q13.2	4680	Hs.73848	163980
205498_at		NM_01446 7.1	0.00000003	Up	2.03879297	sushi-repeat protein	electron transport //	xq21.33- q23	27286	Hs.126782	
204341_at		NM_00647 0.1	0	Up	2.04802937	tripartite motif- containing 16	cytoplasm //				
219956_at	GALNT6	NM_00721 0.2	0.00000002	Up	2.05001443	UDP-N-acetyl-alpha- D- galactosamine:polypeptide N- acetylgalactosaminyl transferase 6 (GalNAc T6)	polypeptide N- acetylgalactosaminyltra nsferase activity //	12q13	11226	Hs.151678	605148
204058_at	ME1	AL049698	0	Up	2.10485729		extended; inferred from electronic annotation; 2.4.1.41; 3.82e-162	6q12	4189	Hs.14732	154250
221841_s_at		BF514079	0.00000918	Up	2.2085237	Kruppel-like factor 4 (gut)	tricarboxylic acid cycle // not recorded NULL			Hs.7934	

FIG. 6M

60/81

Gene list BU AFX_ID	GENE_SYM	GB_ACC	Smoker_NS Smoker_p Value	Up or Down regulation	Fold Change	GENBANK DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
210505_at	ADH7	U07821.1	0.00000089	Up	2.22191181	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	regulation of transcription, DNA- dependent // traceable author statement	4q23-q24	131	Hs.389	600086
209213_at	CBR1	BC002511.1	0.00000429	Up	2.30152926	carbonyl reductase 1	metabolism // inferred from electronic annotation	21q22.13	873	Hs.88778	114830
204059_s_at	ME1	NM_00239 5.2	0	Up	2.4145627	malic enzyme 1, NADP(+)-dependent, cytosolic	tricarboxylic acid cycle // not recorded	6q12	4199	Hs.14732	154250
209386_at	TM4SF1	AI346835	0.00001661	Up	2.61139519	transmembrane 4 superfamily member 1	integral to plasma membrane // inferred from direct assay	3q21-q25	4071	Hs.351316	191155
207469_s_at		NM_00366 2.1	0	Up	2.7389925	Pirin	transcription from Pol II promoter // traceable author statement	xp22.31	8544	Hs.279663	603329
204151_x_at	AKR1C1	NM_00135 3.2	0.00000293	Up	2.8545194	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20- alpha (3-alpha)- hydroxysteroid dehydrogenase)	xenobiotic metabolism // traceable author statement	10p15-p14	1645	Hs.306098	600449
217678_at		AA488687	0.00000146	Up	2.87500492	ESTs	NULL			Hs.284235	
210519_s_at	NQO1	BC000906.1	0	Up	3.07675216		electron transport // inferred from electronic annotation	16q22.1	1728	Hs.80706	125860

FIG. 6N



Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
209160_at	AKR1C3	AB018580.1	0.00000041	Up	3.14061275	aldo-keto reductase family 1, member C3 (3-alpha hydroxysteroid dehydrogenase, type II)	lipid metabolism // traceable author statement	10p15-p14	8644	Hs.78183	603966
216594_x_at		S68290.1	0.00000003	Up	3.37268899		xenobiotic metabolism // traceable author statement				
214385_s_at	MUC5B	A1521646	0.00000003	Up	3.38189534	mucin 5; subtype B, tracheobronchial	protein binding // inferred from physical interaction	11p15	4587	Hs.102482	600770
202831_at	GPX2	NM_002083.1	0	Up	3.42949443	glutathione peroxidase 2 (gastrointestinal)	response to oxidative stress // inferred from electronic annotation	14q24.1	2877	Hs.2704	138319
205328_at	CLDN10	NM_006984.1	0	Up	3.4329732	claudin 10	cell adhesion // experimental evidence	13q31-q34	9071	Hs.26126	
201468_s_at	NQO1	NM_000903.1	0	Up	3.46737099	NAD(P)H dehydrogenase, quinone 1	electron transport // inferred from electronic annotation	16q22.1	1728	Hs.80706	125860
211653_x_at		M33376.1	0.00000001	Up	3.55079646		NULL				
214303_x_at		AW192795	0.00000018	Up	3.62489718	mucin 5, subtypes A and C, tracheobronchial/gastric	protein binding // inferred from physical interaction			Hs.103707	
201467_s_at	NQO1	A1039874	0	Up	4.0084017	NAD(P)H dehydrogenase, quinone 1	electron transport // inferred from electronic annotation	16q22.1	1728	Hs.80706	125860

FIG. 60

62/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
209699_x_at	AKR1C2	U05598.1	0.000000016	Up	4.21436822	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III)	canalicular bile acid transport // traceable author statement	10p15-p14	1646	Hs.201967	600450
209921_at		AB040875. 1	0.000000061	Up	4.86460476	solute carrier family 7, (cationic amino acid transporter, y+ system) member 11	transport // extended:Unknown; aa_permeases; 2e-05			Hs.6682	
217626_at		BF508244	0.000000008	Up	5.28991454	ESTs, Highly similar to DBDD_HUMAN TRANS-1,2- DIHYDROBENZENE- 1,2-DIOL DEHYDROGENASE [H.sapiens]	xenobiotic metabolism // traceable author statement			Hs.300832	
201884_at	CEACAM5	NM_00436 3.1	0.000000048	Up	5.77915081	carcinoembryonic antigen-related cell adhesion molecule 5	integral to plasma membrane // traceable author statement	19q13.1- q13.2	1048	Hs.220529	114890
205623_at	ALDH3A1	NM_00069 1.1	0	Up	6.06762451	aldehyde dehydrogenase 3 family, member A1	aldehyde metabolism // predicted/computed	17p11.2	218	Hs.575	100660

FIG. 6P

63/81

Gene list BUAFFX_ID	GENE_SYM	GB_ACC	Smoker_NS smoker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
202435_s_at	CYP1B1	NM_00010 4.2	0.00000961	Up	7.09658827	cytochrome P450, subfamily I (dioxin- inducible), polypeptide 1 (glaucoma 3, primary infantile)	electron transport // inferred from electronic annotation	2p21	1545	Hs.154654	601771
202436_s_at	CYP1B1	NM_00010 4.2	0.00000296	Up	14.65084839	cytochrome P450, subfamily I (dioxin- inducible), polypeptide 1 (glaucoma 3, primary infantile)	electron transport // inferred from electronic annotation	2p21	1545	Hs.154654	601771
202437_s_at	CYP1B1	NM_00010 4.2	0.00000055	Up	25.18444139	cytochrome P450, subfamily I (dioxin- inducible), polypeptide 1 (glaucoma 3, primary infantile)	electron transport // inferred from electronic annotation	2p21	1545	Hs.154654	601771
Duplicated probe sets											
Probe set	Gene ID	GB_Access ion									
208918_s_at		BC001709. 1									
208919_s_at		BC001709. 1									

FIG. 6Q

Gene list BU.AFFX_ID	GENE_SYM	GB_ACC	Smoker_NS moker_p Value	Up or Down regulation	Fold Change	GENBANK_ DESCRIPTION	Functional category	MAP	LOCUS LINK	UNIGENE	OMIM
202435_s_at	CYP1B1	NM_00010 4.2									
202436_s_at	CYP1B1	NM_00010 4.2									
202437_s_at	CYP1B1	NM_00010 4.2									

FIG. 6R

65/81

Probe Set ID	Gene Title	Gene Symbol	Chromosomal Location	Locus Link
200872_at	S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))	S100A10	1q21	6281
200884_at	creatine kinase, brain	CKB	14q32	1152
201118_at	phosphogluconate dehydrogenase /// phosphogluconate dehydrogenase	PGD	1p36.3-p36.13 /// 1p36.3-p36.13	5226
201234_at	integrin-linked kinase	ILK	11p15.5-p15.4	3611
201266_at	thioredoxin reductase 1	TXNRD1	12q23-q24.1	7296
201435_s at	eukaryotic translation initiation factor 4E	EIF4E	4q21-q25	1977
201467_s at	NAD(P)H dehydrogenase, quinone 1	NQO1	16q22.1	1728
201468_s at	NAD(P)H dehydrogenase, quinone 1	NQO1	16q22.1	1728
201502_s at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	NFKBIA	14q13	4792
201884_at	carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM5	19q13.1-q13.2	1048
202018_s at	lactotransferrin	LTF	3q21-q23	4057
202177_at	growth arrest-specific 6	GAS6	13q34	2621
202303_x at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5	SMARCA5	4q31.1-q31.2	8467
202381_at	a disintegrin and metalloproteinase domain 9 (meltrin gamma)	ADAM9	8p11.22	8754
202435_s at	cytochrome P450, family 1, subfamily B, polypeptide 1	CYP1B1	2p21	1545
202436_s at	cytochrome P450, family 1, subfamily B, polypeptide 1	CYP1B1	2p21	1545

FIG. 7A

202437 s at	cytochrome P450, family 1, subfamily B, polypeptide 1	CYP1B1	2p21	1545
202831 at	glutathione peroxidase 2 (gastrointestinal)	GPX2	14q24.1	2877
202922 at	glutamate-cysteine ligase, catalytic subunit	GCLC	6p12	2729
202923 s at	glutamate-cysteine ligase, catalytic subunit	GCLC	6p12	2729
203240 at	Fc fragment of IgG binding protein	FCGBP	19q13.1	8857
203290 at	major histocompatibility complex, class II, DQ alpha 1	HLA-DQA1	6p21.3	3117
203382 s at	apolipoprotein E	APOE	19q13.2	348
203397 s at	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylglucosaminyltransferase 3 (GalNAc-T3)	GALNT3	2q24-q31	2591
203687 at	chemokine (C-X3-C motif) ligand 1	CX3CL1	16q13	6376
203757 s at	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	CEACAM6	19q13.2	4680
203824 at	transmembrane 4 superfamily member 3	TM4SF3	12q14.1-q21.1	7103
203925 at	glutamate-cysteine ligase, modifier subunit	GCLM	1p22.1	2730
203963 at	carbonic anhydrase XII	CA12	15q22	771
203980 at	fatty acid binding protein 4, adipocyte	FABP4	8q21	2167
203994 s at	chromosome 21 open reading frame 2	C21orf2	21q22.3	755
204041 at	monoamine oxidase B	MAOB	Xp11.23	4129
204058 at	malic enzyme 1, NADP(+)-dependent, cytosolic	ME1	6q12	4199
204059 s at	malic enzyme 1, NADP(+)-dependent, cytosolic	ME1	6q12	4199
204146 at	RAD51-interacting protein	PIR51	12p13.2-p13.1	10635

FIG. 7B

67/81

204151_x_at	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)	AKR1C1	10p15-p14	1645
204326_x_at	metallothionein 1X /// metallothionein 1X	MT1X	16q13 /// 16q13	4501
204379_s_at	fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)	FGFR3	4p16.3	2261
204416_x_at	apolipoprotein C-I	APOC1	19q13.2	341
204420_at	hepatitis delta antigen-interacting protein A	DIPA	11q12.1	11007
204441_s_at	polymerase (DNA-directed), alpha (70kD)	POLA2	11q13.1	23649
204734_at	keratin 15	KRT15	17q21.2	3866
204745_x_at	metallothionein 1G	MT1G	16q13	4495
204755_x_at	hepatic leukemia factor	HLF	17q22	3131
204790_at	MAD, mothers against decapentaplegic homolog 7 (Drosophila)	MADH7	18q21.1	4092
204878_s_at	prostate derived STE20-like kinase PSK /// prostate derived STE20-like kinase PSK	PSK	16p12.1 /// 16p12.1	51677
205048_s_at	---	---	---	---
205244_s_at	solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 3	SLC13A3	20q12-q13.1	64849
205328_at	claudin 10	CLDN10	13q31-q34	9071
205473_at	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B, isoform 1 (Renal tubular acidosis with deafness)	ATP6V1B1	2p13.1	525
205513_at	transcobalamin I (vitamin B12 binding protein, R binder family)	TCN1	11q11-q12	6947

FIG. 7C

68/81

205623 at	aldehyde dehydrogenase 3 family, member A1	ALDH3A1	17p11.2	218
205680 at	matrix metalloproteinase 10 (stromelysin 2)	MMP10	11q22.3	4319
205725 at	secretoglobulin, family 1A, member 1 (uteroglobulin)	SCGB1A1	11q12.3-q13.1	7356
205749 at	cytochrome P450, family 1, subfamily A, polypeptide 1	CYP1A1	15q22-q24	1543
205883 at	zinc finger protein 145 (Kruppel-like, expressed in promyelocytic leukemia)	ZNF145	11q23.1	7704
206461 x at	metallothionein 1H	MT1H	16q13	4496
206561 s at	aldo-keto reductase family 1, member B10 (aldose reductase)	AKR1B10	7q33	57016
206669 at	glutamate decarboxylase 1 (brain, 67kDa)	GAD1	2q31	2571
206998 x at	proline-rich protein BstNI subfamily 4	PRB4	12p13.2	5545
207232 s at	zinc finger DAZ interacting protein 3	DZIP3	3q13.13	9666
207244 x at	cytochrome P450, family 2, subfamily A, polypeptide 6	CYP2A6	19q13.2	1548
207291 at	transmembrane gamma-carboxyglutamic acid protein 4 III transmembrane gamma-carboxyglutamic acid protein 4	TMG4	11p13 /// 11p13	79056
207469 s at	Pirin	PIR	Xp22.31	8544
207605 x at	Kruppel-related zinc finger protein	H-plk	7q11.21	51351
208063 s at	calpain 9 (nCL-4)	CAPN9	1q42.11-q42.3	10753
208327 at	cytochrome P450, family 2, subfamily A, polypeptide 13	CYP2A13	19q13.2	1553
208348 s at	Cas-Br-M (murine) ecotropic retroviral transforming sequence b	CBLB	3q13.11-q13.12	868

FIG. 7D



69/81

208442_s_at	ataxia telangiectasia mutated (includes complementation groups A, C and D)	ATM	11q22-q23	472
208581_x_at	metallothionein 1X	MT1X	16q13	4501
208712_at	cyclin D1 (PRAD1; parathyroid adenomatosis 1)	CCND1	11q13	595
209074_s_at	TU3A protein	TU3A	3p21.1	11170
209101_at	connective tissue growth factor	CTGF	6q23.1	1490
209160_at	aldo-keto reductase family 1, member C3 (3-alpha hydroxysteroid dehydrogenase, type II)	AKR1C3	10p15-p14	8644
209213_at	carbonyl reductase 1	CBR1	21q22.13	873
209351_at	keratin 14 (epidermolysis bullosa simplex; Dowling-Meara, Koebner)	KRT14	17q12-q21	3861
209369_at	annexin A3	ANXA3	4q13-q22	306
209386_at	transmembrane 4 superfamily member 1	TM4SF1	3q21-q25	4071
209387_s_at	transmembrane 4 superfamily member 1	TM4SF1	3q21-q25	4071
209465_x_at	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1)	PTN	7q33-q34	5764
209468_at	low density lipoprotein receptor-related protein 5	LRP5	11q13.4	4041
209699_x_at	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III)	AKR1C2	10p15-p14	1646
209784_s_at	jagged 2	JAG2	14q32	3714
209894_at	leptin receptor	LEPR	1p31	3953
209921_at	solute carrier family 7, (cationic amino acid transporter, y+ system) member 11	SLC7A11	4q28-q32	23657

FIG. 7E

70/81

209987 s at	achaete-scute complex-like 1 (Drosophila)	ASCL1	12q22-q23	429
210064 s at	uroplakin 1B	UPK1B	3q13.3-q21	7348
210128 s at	leukotriene B4 receptor	LTB4R	14q11.2-q12	1241
210179 at	potassium inwardly-rectifying channel, subfamily J, member 13	KCNJ13	2q37	3769
210337 s at	ATP citrate lyase	ACLY	17q12-q21	47
210376 x at	ELK1, member of ETS oncogene family	ELK1	Xp11.2	2002
210402 at	potassium inwardly-rectifying channel, subfamily J, member 1	KCNJ1	11q24	3758
210445 at	fatty acid binding protein 6, ileal (gastrotrypin)	FABP6	5q23-q35	2172
210495 x at	fibronectin 1	FN1	2q34	2335
210505 at	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	ADH7	4q23-q24	131
210517 s at	A kinase (PRKA) anchor protein (gravin) 12	AKAP12	6q24-q25	9590
210519 s at	NAD(P)H dehydrogenase, quinone 1	NQO1	16q22.1	1728
210524 x at	metallothionein 1F (functional)	MT1F	16q13	4494
210705 s at	tripartite motif-containing 5	TRIM5	---	85363
211090 s at	---	---	---	---
211295 x at	cytochrome P450, family 2, subfamily A, polypeptide 6	CYP2A6	19q13.2	1548

FIG. 7F

211456_x_at	---					
211564_s_at	LIM domain protein	RIL	5q31.1			8572
211653_x_at	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III) /// aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III)	AKR1C2	10p15-p14 /// 10p15-p14			1646
211657_at	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen) /// carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	CEACAM6	19q13.2 /// 19q13.2			4680
211689_s_at	transmembrane protease, serine 2 /// transmembrane protease, serine 2	TMPRSS2	21q22.3 /// 21q22.3			7113
211796_s_at	---	---	---			---
211885_x_at	fucosyltransferase 6 (alpha (1,3) fucosyltransferase)	FUT6	19p13.3			2528
212026_s_at	likely ortholog of mouse exocyst component protein 70 kDa homolog (S. cerevisiae) Exo70: exocyst component protein 70 kDa homolog (S. cerevisiae)	EXO70	17q25.3			23265
212223_at	iduronate 2-sulfatase (Hunter syndrome)	IDS	Xq28			3423
212805_at	KIAA0367 protein	KIAA0367	9q21.31			23273
213182_x_at	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	CDKN1C	11p15.5			1028
213432_at	mucin 5, subtypes A and C, tracheobronchial/gastric	MUC5AC	11p15			4586
213530_at	RAB3 GTPase-ACTIVATING PROTEIN	RAB3GAP	2q21.3			22930

FIG. 7G

72/81

213629 x at	metallothionein 1F (functional)	MT1F	16q13	4494
213796 at	small proline-rich protein 1A	SPRR1A	1q21-q22	6698
213859 x at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5	SMARCA5	4q31.1-q31.2	8467
214164 x at	hypothetical protein FLJ20151	FLJ20151	15q22.1	54837
214186 s at	Homo sapiens (clone 3.8-1) MHC class I mRNA fragment.	---	---	---
214303 x at	mucin 5, subtypes A and C, tracheobronchial/gastric	MUC5AC	11p15	4586
214320 x at	---	---	---	---
214385 s at	mucin 5, subtypes A and C, tracheobronchial/gastric	MUC5AC	11p15	4586
214830 at	solute carrier family 38, member 6	SLC38A6	14q23.1	145389
215363 x at	folate hydrolase (prostate-specific membrane antigen) 1	FOLH1	11p11.2	2346
215655 at	glutamate receptor, ionotropic, kainate 2	GRIK2	6q16.3-q21	2898
215867 x at	adaptor-related protein complex 1, gamma 1 subunit	AP1G1	16q23	164
215913 s at	PTB domain adaptor protein CED-6	CED-6	2q32.3-q33	51454
216284 at	Homo sapiens cDNA FLJ14061 fis, clone HEMBB1000749.	---	---	---
216346 at	---	---	---	---
216442 x at	fibronectin 1	FN1	2q34	2335

FIG. 7H

216594 x at	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)	AKR1C1	10p15-p14	1645
217022 s at	hypothetical protein MGC27165	MGC27165	14q32.33	283650
217059 at	mucin 7, salivary	MUC7	4q13-q21	4589
217129 at	---	---	---	---
217165 x at	---	---	---	---
217188 s at	---	---	---	---
217294 s at	enolase 1, (alpha)	ENO1	1p36.3-p36.2	2023
217382 at	---	---	---	---
217414 x at	---	---	---	---
217524 x at	phosphodiesterase 6A, cGMP-specific, rod, alpha	PDE6A	5q31.2-q34	5145
217626 at	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III)	AKR1C2	10p15-p14	1646
217635 s at	polymerase (DNA directed), gamma	POLG	15q25	5428
217671 at	Homo sapiens transcribed sequence with weak similarity to protein prf:2109260A (H.sapiens) 2109260A B cell growth factor [Homo sapiens]	---	---	---
217678 at	solute carrier family 7, (cationic amino acid transporter, y+ system) member 11	SLC7A11	4q28-q32	23657
217897 at	FXD domain containing ion transport regulator 6	FXD6	11q23.3	53826

FIG. 7I

217979 at	transmembrane 4 superfamily member 13	TM4SF13	7p21.2	27075
218225 at	likely ortholog of mouse signaling intermediate in Toll pathway- evolutionarily conserved	SITPEC	19p13.2	51295
218313 s at	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N- acetyl-galactosaminyltransferase 7 (GalNAc-T7)	GALNT7	4q31.1	51809
218418 s at	KIAA1518 protein	KIAA1518	19p13.2	25959
218566 s at	cysteine and histidine-rich domain (CHORD)-containing, zinc binding protein 1	CHORDC1	11q14.3	26973
218579 s at	DEAH (Asp-Glu-Ala-His) box polypeptide 35	DHX35	20q11.22-q12.1	60625
218700 s at	RAB7, member RAS oncogene family-like 1	RAB7L1	1q32	8934
218938 at	hypothetical protein MGC11279	MGC11279	10q24.32	79176
218990 s at	small proline-rich protein 3	SPRR3	1q21-q22	6707
219040 at	hypothetical protein FLJ22021	FLJ22021	16p13.3	79585
219359 at	hypothetical protein FLJ22635	FLJ22635	11p15.5	80162
219404 at	EPS8-like 3	EPS8L3	1p13.2	79574
219457 s at	Ras and Rab interactor 3	RIN3	14q32.13	79890
219944 at	hypothetical protein FLJ21069	FLJ21069	2p23.3	79745
219956 at	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N- acetyl-galactosaminyltransferase 6 (GalNAc-T6)	GALNT6	12q13	11226
220013 at	abhydrolase domain containing 9	ABHD9	19p13.13	79852
220262 s at	EGF-like-domain, multiple 9	EGFL9	6p21.1	65989

FIG. 7J

75/81

220266 s at	Kruppel-like factor 4 (gut)	KLF4	9q31	9314
220389 at	hypothetical protein FLJ23514	FLJ23514	11q14.1	60494
220432 s at	cytochrome P450, family 39, subfamily A, polypeptide 1	CYP39A1	6p21.1-p11.2	51302
220542 s at	palate, lung and nasal-epithelium carcinoma associated	PLUNC	20q11.2	51297
220562 at	hypothetical protein FLJ20359	FLJ20359	7p22.3	54905
220798 x at	hypothetical protein FLJ11535	FLJ11535	19p13.3	79948
221079 s at	methylintransferase like 2	METTL2	7q32.2	55798
221359 at	glial cell derived neurotrophic factor	GDNF	5p13.1-p12	2668
221511 x at	cell cycle progression 8 protein	CPR8	15q21.1	9236
221568 s at	lin-7 homolog C (C. elegans)	LIN7C	11p14	55327
221577 x at	prostate differentiation factor	PLAB	19p13.1-13.2	9518
221606 s at	nucleosomal binding protein 1	NSBP1	Xq13.3	79366
221781 s at	ER-resident protein ERdj5	ERdj5	2q32.1	54431
221841 s at	Kruppel-like factor 4 (gut)	KLF4	9q31	9314
221942 s at	guanylate cyclase 1, soluble, alpha 3	GUCY1A3	4q31.3-q33	2982
222016 s at	zinc finger protein 323	ZNF323	---	64288

FIG. 7K

76/81

222071_s at	solute carrier organic anion transporter family, member 4C1	SLCO4C1	5q21.2	353189
37953_s at	amiloride-sensitive cation channel 2, neuronal	ACCN2	12q12	41
40093_at	Lutheran blood group (Auberger b antigen included)	LU	19q13.2	4059
823_at	chemokine (C-X3-C motif) ligand 1	CX3CL1	16q13	6376

FIG. 7L



77/81

Probe Set ID	Gene Title	Gene Symbol	Chromosomal Location	Locus Link
200772_x_at	prothymosin, alpha (gene sequence 28)	PTMA	2q35-q36	5757
201481_s_at	phosphorylase, glycogen; brain	PYGB	20p11.2-p11.1	5834
201486_at	reticulocalbin 2, EF-hand calcium binding domain	RCN2	15q23	5955
202018_s_at	lactotransferrin	LTF	3q21-q23	4057
202163_s_at	CCR4-NOT transcription complex, subunit 8	CNOT8	5q31-q33	9337
202335_s_at	ubiquitin-conjugating enzyme E2B (RAD6 homolog)	UBE2B	5q23-q31	7320
202445_s_at	Notch homolog 2 (Drosophila)	NOTCH2	1p13-p11	4853
202733_at	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide II	P4HA2	5q31	8974
202905_x_at	Nijmegen breakage syndrome 1 (nibrin)	NBS1	8q21	4683
202976_s_at	Rho-related BTB domain containing 3	RHOBTB3	5q15	22836
203049_s_at	KIAA0372	KIAA0372	5q15	9652
203284_s_at	heparan sulfate 2-O-sulfotransferase 1	HS2ST1	1p31.1-p22.1	9653
203301_s_at	cyclin D binding myb-like transcription factor 1	DMTF1	7q21	9988
203551_s_at	COX11 homolog, cytochrome c oxidase assembly protein (yeast)	COX11	17q22	1353
203695_s_at	deafness, autosomal dominant 5	DFNA5	7p15	1687
204018_x_at	hemoglobin, alpha 2	HBA2	16p13.3	3040
204135_at	downregulated-in ovarian cancer 1	DOC1	3q12.3	11259
204410_at	eukaryotic translation initiation factor 1A, Y-linked	EIF1AY	Yq11.222	9086
204419_x_at	hemoglobin, gamma G	HBG2	11p15.5	3048
204451_at	frizzled homolog 1 (Drosophila)	FZD1	7q21	8321
204521_at	protein predicted by clone 23733	HSU79274	12q24.13	29902
204790_at	MAD, mothers against decapentaplegic homolog 7 (Drosophila)	MADH7	18q21.1	4092
204832_s_at	bone morphogenetic protein receptor, type IA	BMPRI1A	10q22.3	657
204848_x_at	hemoglobin, gamma A	HBG1	11p15.5	3047
205034_at	cyclin E2	CCNE2	8q22.1	9134
205048_s_at	—	—	—	—
205091_x_at	RecQ protein-like (DNA helicase Q1-like)	RECQL	12p12	5965

FIG. 8A

78/81

205100_at	glutamine-fructose-6-phosphate transaminase 2	GFPT2	5q34-q35	9945
205221_at	homogentisate 1,2-dioxygenase (homogentisate oxidase)	HGD	3q21-q23	3081
205458_at	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)	MC1R	16q24.3	4157
205684_s_at	hypothetical protein FLJ20686	FLJ20686	9p21.3	55667
205822_s_at	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	HMGCS1	5p14-p13	3157
205857_at	solute carrier family 18 (vesicular monoamine), member 2	SLC18A2	10q25	6571
205911_at	parathyroid hormone receptor 1	PTH1R	3p22-p21.1	5745
205954_at	retinoid X receptor, gamma	RXRG	1q22-q23	6258
205989_s_at	myelin oligodendrocyte glycoprotein	MOG	6p22-p21.3	4340
206242_at	transmembrane 4 superfamily member 5	TM4SF5	17p13.3	9032
206279_at	protein kinase, Y-linked	PRKY	Yp11.2	5616
206834_at	hemoglobin, delta	HBD	11p15.5	3045
207158_at	apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1	APOBEC1	12p13.1	339
207308_at	solute carrier organic anion transporter family, member 1A2	SLCO1A2	12p12	6579
207345_at	folistatin	FST	5q11.2	10468
207404_s_at	5-hydroxytryptamine (serotonin) receptor 1E	HTR1E	6q14-q15	3354
207777_s_at	SP140 nuclear body protein	SP140	2q37.1	11262
207786_at	cytochrome P450, family 2, R1	CYP2R1	11p15.2	120227
208152_s_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	DDX21	10q21 /// 10q21	9188
208292_at	bone morphogenetic protein 10	BMP10	2p13.3	27302
208410_x_at	amelogenin (amelogenesis imperfecta 1, X-linked)	AMELX	Xp22.31-p22.1	265
208525_s_at	olfactory receptor, family 2, subfamily F, member 1	OR2F1	7q35	26211
208541_x_at	transcription factor A, mitochondrial	TFAM	10q21	7019
208551_at	histone 1, H4g	HIST1H4G	6p21.3	8369
208896_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18	DDX18	2q14.2	8886
209098_s_at	jagged 1 (Alagille syndrome)	JAG1	20p12.1-p11.23	182
209458_x_at	hemoglobin, alpha 2	HBA2	16p13.3	3040
209916_at	dehydrogenase E1 and transketolase domain containing 1	DHDKD1	10p14	55526

FIG. 8B

79/81

210552_s_at	Ral guanine nucleotide exchange factor RalGPS1A	RALGPS1A	9q34.13	9649
210596_at	implantation-associated protein	DKFZp564K142	Xq13.3	84061
210690_at	killer cell lectin-like receptor subfamily C, member 4	KLRC4	12p13.2-p12.3	8302
210818_s_at	BTB and CNC homology 1, basic leucine zipper transcription factor 1	BACH1	21q22.11	571
210868_s_at	ELOVL family member 6, elongation of long chain fatty acids (FEN1/Elo2, SUR4/Elo3-like, yeast)	ELOVL6	4q25	79071
211241_at	Human lipocortin (LIP) 2 pseudogene mRNA, complete cds-like region.	—	—	—
211699_x_at	hemoglobin, alpha 2 /// hemoglobin, alpha 2	HBA2	16p13.3 /// 16p13.3	3040
211921_x_at	Homo sapiens fetal thymus prothymosin alpha mRNA, complete cds /// Homo sapiens fetal thymus prothymosin alpha mRNA, complete cds	—	///	—
211965_at	zinc finger protein 36, C3H type-like 1	ZFP36L1	14q22-q24	677
211967_at	pro-oncogene receptor inducing membrane injury gene	PORIMIN	11q22.1	114908
212220_at	proteasome (prosome, macropain) activator subunit 4	PSME4	2p16.3	23198
212277_at	myotubularin related protein 4	MTMR4	17q22-q23	9110
212628_at	protein kinase C-like 2	PRKCL2	1p22.2	5586
212942_s_at	KIAA1199 protein	KIAA1199	15q	57214
213402_at	hypothetical protein LOC126208	LOC126208	19q13.43	126208
213439_x_at	RaP2 Interacting protein 8	RPIP8	17q21.31	10900
214153_at	ELOVL family member 5, elongation of long chain fatty acids (FEN1/Elo2, SUR4/Elo3-like, yeast)	ELOVL5	6p21.1-p12.1	60481
214442_s_at	Msx-Interacting-zinc finger	MIZ1	18q21.1	9063
214458_at	microtubule-interacting protein that associates with TRAF3	MIP-T3	2q37.3	26146
214599_at	involucrin	IVL	1q21	3713
214684_at	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)	MEF2A	15q26	4205
214765_s_at	N-acylsphingosine amidohydrolase (acid ceramidase)-like	ASAH1	4q21.1	27163
214847_s_at	chromosome 6 open reading frame 9	C6orf9	6p21.3	63940
214850_at	SMA5	SMA5	5q13	11042
214871_x_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	SMARCE1	17q21.2	6605

FIG. 8C

214931_s_at	SFRS protein kinase 2	SRPK2	7q22-q31.1	6733
214975_s_at	myotubularin related protein 1	MTMR1	Xq28	8776
215009_s_at	yeast Sec31p homolog	KIAA0905	4q21.3	22872
215174_at	Homo sapiens cDNA FLJ27001 fis, clone SLV04836	—	1q24.1	375037
215220_s_at	translocated promoter region (to activated MET oncogene)	TPR	1q25	7175
215387_x_at	Homo sapiens cDNA FLJ11443 fis, clone HEMBA1001330.	—	—	—
215586_at	Homo sapiens cDNA FLJ14111 fis, clone MAMMA1001630.	—	—	—
215609_at	Homo sapiens cDNA FLJ13833 fis, clone THYRO1000676.	—	—	—
216184_s_at	regulating synaptic membrane exocytosis 1	RIMS1	6q12-q13	22999
216374_at	—	—	—	—
216540_at	TCR V alpha 14.1/J alpha 32/C alpha =melanoma antigen-specific T-cell receptor alpha chain {V-J-C junction, alternatively spliced, clone TIL C10-1} [human, CD8+ tumor-infiltrating lymphocytes, mRNA Partial, 108 nt]	—	—	—
216556_x_at	—	—	—	—
216933_x_at	adenomatosis polyposis coli	APC	5q21-q22	324
217097_s_at	—	—	—	—
217313_at	—	—	—	—
217573_at	glutamate receptor, ionotropic, N-methyl D-aspartate 2C	GRIN2C	17q25	2905
217578_at	Homo sapiens transcribed sequence with moderate similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]	—	—	—
217881_s_at	cell division cycle 27	CDC27	17q12-17q23.2	996
218433_at	pantothenate kinase 3	PANK3	5q34	79646
218508_at	transcription factor SMIF	HSA275986	3p21.31	55802
218802_at	hypothetical protein FLJ20647	FLJ20647	4q25	55013
218994_s_at	hypothetical protein FLJ13195 similar to stromal antigen 3	FLJ13195	7p11.2-q11.2	64940
219029_at	hypothetical protein FLJ21657	FLJ21657	5p12	64417
219099_at	chromosome 12 open reading frame 5	C12orf5	12p13.3	57103
219148_at	T-LAK cell-originated protein kinase	TOPK	8p21.2	55872
219242_at	hypothetical protein FLJ13386	FLJ13386	3q22.1	80254
219363_s_at	CGI-12 protein	CGI-12	8q22.1	51001

FIG. 8D

219433_at	BCL6 co-repressor	BCOR	Xp11.4	54880
219470_x_at	hypothetical protein FLJ10895	FLJ10895	10pter-q26.12	54619
219588_s_at	hypothetical protein FLJ20311	FLJ20311	7q36.3	54892
219750_at	hypothetical protein FLJ11155	FLJ11155	4q32.1	55314
219920_s_at	GDP-mannose pyrophosphorylase B	GMPPB	3p21.31	29925
220172_at	hypothetical protein FLJ13096	FLJ13096	2q31.1	80067
220334_at	regulator of G-protein signalling 17	RGS17	6q25-q26	26575
220422_at	ubiquitin 3	UBQLN3	11p15	50613
220459_at	MCM3 minichromosome maintenance deficient 3 (S. cerevisiae) associated protein, antisense	MCM3APAS	21q22.3	114044
220856_x_at	---	---	---	---
220934_s_at	---	---	---	---
221045_s_at	period homolog 3 (Drosophila)	PER3	1p36.23	8863
221451_s_at	olfactory receptor, family 2, subfamily W, member 1 /// olfactory receptor, family 2, subfamily W, member 1	OR2W1	6p21.31-21.33 6p21.31-21.33	26692
221615_at	peptidylprolyl isomerase E (cyclophilin E)	PPIE	1p32	10450
221826_at	similar to RIKEN cDNA 2610307121	LOC90806	1q32.3	90806
222006_at	fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)	FGFR3	4p16.3	2261
222071_s_at	solute carrier organic anion transporter family, member 4C1	SLCO4C1	5q21.2	353189
222107_x_at	leucine zipper, putative tumor suppressor 1	LZTS1	8p22	11178
51228_at	Homo sapiens similar to RNA binding motif protein 12 (LOC375676), mRNA	---	8q22.1	375676
54051_at	PBX/knotted 1 homeobox 1	PKNOX1	21q22.3	5316

FIG. 8E

**THIS PAGE BLANK (USPTO)**